DOCUMENT RESUME

ED 258 596 IR 051 210

| | |
|---|---|
| AUTHOR | Naugler, Harold |
| TITLE | The Archival Appraisal of Machine-Readable Records: A RAMP Study with Guidelines. |
| INSTITUTION | United Nations Educational, Scientific and Cultural Organization, Paris (France). General Information Programme. |
| REPORT NO | PGI-84/WS/27 |
| PUB DATE | Nov 84 |
| NOTE | 171p.; For related documents, see IR 051 188-189. |
| PUB TYPE | Guides - General (050) -- Viewpoints (120) -- Tests/Evaluation Instruments (160) |
| EDRS PRICE | MF01 Plus Postage. PC Not Available from EDRS. |
| DESCRIPTORS | *Archives; Computers; Computer Software; Confidentiality; Developed Nations; Developing Nations; Evaluation Criteria; *Evaluation Methods; Guidelines; *Library Material Selection; *Policy; Preservation; *Records (Forms) |
| IDENTIFIERS | *Archival Value; *Machine Readable Data; Records Management |

ABSTRACT

Prepared for Unesco's Records and Archive Management Programme (RAMP), this study is designed to provide archivists and other interested information professionals with an introduction to machine-readable records and to provide guidelines for the appraisal of their archival value. The study assumes no prior knowledge of machine-readable records and should be equally useful to archivists in industrialized and developing countries. The guidelines are based on the policies and practices of those countries with the most experience in this field. This report contains: (1) an introductory discussion of the advantages of computer use in information management and disadvantages of older business systems; (2) background information on computer systems and machine-readable records; (3) procedures for managing machine-readable records; (4) guidelines and criteria for the appraisal of both the content and the technical attributes of machine-readable records; (5) descriptions of two processes for anonymizing personal data to protect confidentiality in machine-readable records; (6) discussion of issues involved in transborder data flow; and (7) summary conclusions in the form of recommended policies and practices, or guidelines. Appendices include appraisal forms and procedures developed by the Public Record Office (United Kingdom), the National Archives and Records Service (U.S. General Services Administration), and the Machine Readable Archives Division of the Public Archives of Canada, and a glossary of terms. A bibliography is also provided. (BBM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED258596

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PGI-84/WS/27

# The Archival appraisal of machine-readable records: a RAMP study with guidelines

BEST COPY AVAILABLE

Reproduced by permission of
Unesco

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
C. Coudert Schklowski
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR051210

General Information Programme and UNISIST

United Nations Educational
Scientific and Cultural Organization

Paris, 1984

Original: English

PGI-84/WS/27
PARIS, November 1984

# THE ARCHIVAL APPRAISAL CF MACHINE-READABLE RECORDS: A RAMP STUDY WITH GUIDELINES

prepared by

Harold Naugler

General Information Programme and UNISIST

United Nations Educational,
Scientific and Cultural Organization

Recommended catalogue entry:

Naugler, Harold.

The Archival appraisal of machine-readable records: a RAMP study with guidelines / prepared by Harold Naugler /for the/ General Information Programme and UNISIST. - Paris: Unesco, 1984. - 161 p.; 30 cm. - (PGI-84/WS/27)

I - Title

II - Unesco General Information Programme and UNISIST

III - Records and Archives Management Programme (RAMP)

© Unesco, 1984

## PREFACE

The Division of the General Information Programme of Unesco in order to better meet the needs of Member States, particularly developing countries, in the specialized areas of records management and archives administration, has developed a long-term Records and Archives Management Programme - RAMP.

The basic elements of the RAMP programme reflect the overall themes of the General Information Programme. RAMP thus includes projects, studies and other activities intended to:

1. Promote the formulation of information policies and plans (national, regional and international).
2. Promote and disseminate methods, norms and standards for information handling.
3. Contribute to the development of information infrastructures.
4. Contribute to the development of specialized information systems in the fields of education, culture and communication, and the natural and social sciences.
5. Promote the training and education of specialists in and users of information.

The purpose of this study, which was prepared under contract with the International Council on Archives, is to provide archivists and other interested information professionals with an introduction to machine-readable records and to provide guidelines for the appraisal of their archival value. The study assumes no prior knowledge of machine-readable records and should be equally useful to archivists in industrialized as well as in developing countries. The guidelines are based upon the policies and practices of those countries that have had the most experience in this field. The study also includes model survey and appraisal forms and a useful glossary of specialized terms.

Comments and suggestions regarding the study are welcomed and should be addressed to the Division of the General Information Programme, UNESCO, 7 place de Fontenoy, 75700 Paris, Other studies prepared under the RAMP programme may also be obtained at the same address.

CONTENTS


| | Page |
|---|---|
| INTRODUCTION | 1 |
| I. BACKGROUND TO COMPUTER SYSTEMS AND MACHINE-READABLE RECORDS | 6 |
| II. MANAGING MACHINE-READABLE RECORDS | 19 |
| III. APPRAISAL OF MACHINE-READABLE RECORDS: PART I | 37 |
| IV. APPRAISAL OF MACHINE-READABLE RECORDS: PART II | 57 |
| V. ADDITIONAL ISSUES TO BE CONSIDERED | 83 |
| VI. GUIDELINES | 97 |
| APPENDICES | |
| A. APPRAISAL FORMS AND PROCEDURES | |
| A.1 SELECTION CRITERIA DEVELOPED BY THE PUBLIC RECORD OFFICE AND AN ASSESSMENT OF DEPARTMENTAL ADP FILES | 103 |
| A.2 EXTRACT FROM THE "DISPOSITION OF FEDERAL RECORDS," A RECORDS MANAGEMENT HANDBOOK DEVELOPED BY THE NATIONAL ARCHIVES AND RECORDS SERVICE, GENERAL SERVICES ADMINISTRATION | 116 |
| A.3 APPRAISAL FORM AND RELATED PROCEDURES DEVELOPED BY THE MACHINE READABLE ARCHIVES DIVISION, PUBLIC ARCHIVES OF CANADA | 134 |
| B. GLOSSARY OF TERMS | 142 |
| BIBLIOGRAPHY | 158 |

## INTRODUCTION

> No other development since the invention of movable type has had as great an effect on the production, dissemination, storage, and use of information as has that of the electronic computer, and this development has only been in process for about thirty-five years. Compared to the other great inventions in information communication (writing, begun about 5,000 years ago; the alphabet, developed some 3,000 years ago; movable type, invented about 700 years ago), the computer is only in its infancy. (1)

What is it that makes the modern electronic computer such a powerful tool in the world today? First of all, electronic computers operate at speeds which are hard to imagine. The time required for the internal operation is measured in nanoseconds (that is, .000 000 001 seconds). A corollary of this speed of work is the volume of work which a computer can do. Examples abound in both industrial and scientific applications where computers are being used to solve problems which would have been insoluable by any practical means because of the sheer volume of calculation involved.

A second characteristic of electronic computers is the consistency with which they carry out their instructions. Machine errors are almost unknown, and because of comprehensive error detection systems they seldom lead to inaccurate results. Most of the errors which are reported with such glee in the press are in fact the result of human error rather than the fault of the machine. If the information fed into the computer is valid and the programs are sound, the machine can be relied upon to produce the results that are required.

A third invaluable characteristic of electronic information processing systems is their great storage capacity. Modern computers can store vast amounts of information in a relatively small space, and in such a way that it can be retrieved and used very rapidly. This great storage capacity is particularly advantageous in applications such as calculation of census data, where very detailed information can be processed in a relatively simple way.

A further advantage of the modern computer is its versatility. For example, the same machine can be programmed to help a company's accountant produce a payroll, help the sales manager analyse a market research report, and assist the company's architects and engineers design a new building.

Another important aspect is the fact that the programming and processing tasks are independent of each other. The machine can be working on any one of a variety of tasks while the computer personnel are preparing a program for yet another piece of work. The machine has the ability to accept detailed instructions and to store these in a high speed internal memory unit. It then has almost immediate access to these instructions, and is not dependent upon an operator feeding in instructions as the work progresses.

At the same time that electronic computers have been developing, the older business systems have been running into difficulties. Some of the problems which these older methods of work have had to face are outlined below.

(1) A growing volume of paperwork. As business and government have become more and more complex, more and more records and reports have been needed. Each organization has needed to keep more detailed information on all aspects of its operations.

(2) Increased costs. As the standard of living has risen, so has the cost of employing labour. This has forced all types of organizations to consider automatic methods of processing information.

(3) Shortage of personnel. Again as society has changed, specialization has increased. A better educated population has led to a decline in the number of workers available for routine processing work.

(4) Elimination of error. In this increasingly complex age, it becomes more and more important that we do not make errors. On a flight to the moon, for example, it is imperative that the results of all calculations be reliable, and that no human transcription errors be made.

(5) The need for rapid decisions. Modern management wants to know what happens as soon as it happens so that the managers can make sensible decisions at the earliest possible moment. In this way, potential business will not be lost because of ignorance on the part of management.

Considering the technical ability of modern computers and the difficulties facing the older methods of information processing, it is not surprising that computers have been introduced into laboratories, business, and government offices throughout most countries. In part these machines are replacing older systems, while in part they are being used to undertake work which could not previously be done.

How have archivists responded to the confluence of converging computer/communication technology, new legislative and management initiatives, the rapid growth in the use of computers, and the explosive growth in the volume of information in machine-readable form? It was events of this nature which led, at least in part, to consideration of the implications of computers by the Fifth International Congress on Archives in 1964 and a year later at the Ninth Meeting of the Archives Round Table. However, at that time few ICA members foresaw the possibility of accessioning machine-readable records. Seven years later, in 1971, at the Thirteenth Meeting of the Round Table, data processing applications and their implications in archives were examined. It was as a result of the report that the Ad Hoc Working Party on the Implications of ADP in Archives was established by the ICA in 1972. The Working Group was the predecessor of the existing Automation Committee of the International Council on Archives. "The deliberations of the Working Party and later, of the Committee led to exchanges of views with regard to

the use of computers for managing archives and the problems of appraising machine-readable records." (2) It was around this same time that a number of national repositories - in Canada, Sweden, the United Kingdom, and the United States - began preparing for the scheduling of machine-readable records and for the acquisition of those appraised as having long-term value.

However, concern for the preservation and use of machine-readable records was not, and is still not, confined to traditional archivists. In 1973 a new international organization was established known as the International Association for Social Science Information Service and Technology (IASSIST). Membership in the Association consisted basically of three groups: the creators and disseminators of machine-readable data, data archivists and data librarians, as well as the users, particularly social scientists, of such data. The data archivists and librarians were representatives of social science data archives which were being established at academic institutions throughout many countries. (3) Although data archivists and librarians do not always have the same background and training as their traditional archival counterparts, both share many of the same concerns with respect to the management of machine-readable records. One particular area in which IASSIST members have provided considerable leadership is in the cataloguing and description of machine-readable data files. (4)

It is interesting to note that much of the early literature written concerning machine-readable records dealt with the crucial question of appraisal. (5) Indeed, this continues to be a topic of considerable interest, discussion, and re-evaluation among archivists who have been dealing with machine-readable records for over a decade. It is therefore most timely that the Division of the General Information Program of UNESCO and the International Council on Archives have agreed to the joint sponsorship of this particular study. Dr. Frank B. Evans, of UNESCO, and Dr. Charles Kecseméti, Executive Secretary of the ICA, were instrumental in initiating the project and established the framework within which the study was to take place. The assignment was the result of a contract between these two organizations.

The author would like to express his appreciation to the members of the ICA Automation Committee who have provided both assistance and advice in reviewing the outline of the study and the first draft. In many respects, this is a collaborative study, inasmuch as the author has drawn heavily upon the work of others in this important field of archival science. A number of these must be acknowledged individually. Charles M. Dollar, the former Director of the Machine-Readable Archives Branch of the National Archives and Records Service in Washington, and Thomas E. Brown, a former senior archivist in the Machine-Readable Archives Branch, have both undertaken a considerable degree of analysis and research in this area. Dr. Dollar presented a paper on the subject at the 1977 annual meeting of the Society of American Archivists (SAA), which was later published in The American Archivist. The basic approach outlined by Dr. Dollar has had widespread acceptance within the archival community, particularly among those who deal with machine-readable records on a day-to-day basis. Dr. Brown expanded on these basic concepts and developed an appraisal workbook which he has used

in workshops presented over the past five years at annual meetings of the SAA.

Throughout 1979/1980, two line managers in the Machine Readable Archives Division of the Public Archives of Canada (PAC), Katharine Gavrel and John McDonald, devoted a considerable amount of time to the development of fairly detailed appraisal guidelines and procedures. Although the drafters of the guidelines drew upon the earlier work undertaken by Dollar and Brown, they went much further in their analysis and approach. The guidelines are now used by staff archivists in the PAC's Machine Readable Archives Division in their appraisal of machine-readable records. The author has drawn extensively from these guidelines in preparing chapters three and four of this study.

Archivists who manage machine-readable records on a full-time basis quickly recognize that procedures which are developed one year may require partial or complete revision in two or three years. This is often necessary in order to keep pace with the many and frequent changes in the computer industry itself. Not only is this the case for the accessioning, processing, and preservation of machine-readable records, but it is also true for the appraisal function. For example, as machine-readable records become admissible as evidence in courts of law throughout various countries, the appraisal of machine-readable records from a legal point of view will become far more important than it is at the present time. As more and more textual information becomes digitized or machine-readable, it will also be necessary to reassess the evidential value of machine-readable records. In other words, the author does not consider the approaches outlined in this study as in any way definitive. While every attempt has been made to reflect the "current state of the art" with respect to the appraisal of machine-readable records, it must be recognized that developments will occur over the years which will necessitate their reassessment and possible revision. The approach should therefore not be interpreted as definitive, but rather as a guideline to archivists who manage machine-readable records.

## NOTES

1. H. Thomas Hickerson, Archives and Manuscripts: An Introduction to Automated Access. Basic Manual Series, Society of American Archivists, Chicago, 1981, page 11.

2. Meyer H. Fishbein, Guidelines For Administering Machine-Readable Archives. Committee on Automation, International Council on Archives, Washington, D.C., November 1980, page 7. This particular publication is an excellent example of the work of the Automation Committee over the years, and particularly some of its members, in addressing problems associated with the archival management of machine-readable records. Committee members have also devoted a great deal of time and attention to the use of computer systems in archives. See, for example, A. Arad and M.E. Olsen, An Introduction to Archival Automation. Committee on Automation, International Council on Archives, Koblenz, Federal Republic of Germany, January 1981. The Committee also produces a journal, ADPA, which contains articles, etc. on both automation in archives and the management of machine-readable records.

3. For an explanation of the reasons for the establishment of such archives, particularly in the United States, and the various functions performed in such institutions, see C. Geda, "Social Science Data Archives," The American Archivist, Volume 42, Number 2, April 1979, pages 158-166.

4. See, for example, the manual written by Sue A. Dodd, Cataloging Machine-Readable Data Files. American Library Association, Chicago, 1982.

5. Meyer H. Fishbein, "Appraising Information in Machine Language Form," The American Archivist, Volume 35, Number 1, January 1972, pages 35-43; L. Bell, The Archival Implications of Machine-Readable Records. Washington, D.C.: VIII International Congress on Archives, 1976; Charles M. Dollar, "Appraising Machine-Readable Records," The American Archivist, Volume 41, Number 4, October 1978, pages 23-30; C.L. Geda, C.W. Austin, and F.X. Blouin, Jr. (eds.), Proceedings of a Conference on Archival Management of Machine-Readable Records, Held at the Bentley Library, the University of Michigan, February 1979. Society of American Archivists, Chicago, 1979.

## Chapter I

## BACKGROUND TO COMPUTER SYSTEMS AND MACHINE-READABLE RECORDS

### BASIC COMPUTER CONCEPTS

1.1 There are two groups of archivists who are interested in computer concepts: those who wish to use computers for intellectual and bibliographic control of archival materials; and those who are interested in the appraisal and accessioning of machine-readable records. The primary purpose of this section of the study is to explain basic terminology associated with data processing and to provide a general overview of how a computer system operates.

### Overview of a Computer System

1.2 A computer system consists of data coming into the system, being processed by it, and leaving the system. Hardware, or physical equipment is used to receive input, process data, and produce output. It is software or instructions which tell the hardware how to process the data.

1.3 Hardware is the physical equipment of a computer system. It is often divided into two types: the Central Processing Unit (CPU) and peripheral equipment. Processing in a computer is carried out in the main memory of the machine which is under the control of the Central Processing Unit. The CPU is the controlling centre of the entire computer system. On the other hand, peripheral equipment is hardware, other than the CPU, which is used to enter data, communicate with the CPU, or produce output. It includes such things as keypunch machines, card readers, tape and disk drives, remote job entry terminals, as well as character and line printers. It is important to note, however, that word processors and mini-computers make the distinction between the CPU and peripheral equipment somewhat obsolete. These devices are self-contained units where the facilities for data entry, processing, and output are all in one place.

1.4 Software consists of instructions (or computer programs) which operate the physical equipment (or hardware) and manipulate the data. It includes "master control" programs which always reside in the computer's main memory and are used universally in all operations at a single computer installation. These programs are called systems software or the computer's operating system. The other type of software is applications software. These programs, usually written in a programming language, instruct the computer to follow a precise sequence of steps in order to manipulate data and produce some desired result.

1.5 Data are representations of facts or concepts that can be communicated by human or automated means. While data can be made up of words, numbers, or special symbols like punctuation marks, in a computer system the letters, numbers, and symbols that constitute data must be represented according to a very precise set of rules.

**Data Representation**

1.6 In order to understand how a computer system operates and how the components of hardware, software, and data fit together, it is necessary to understand the different methods of representing numbers, letters, and special symbols.

1.7 When we communicate, we use what are called natural languages. In our written language, we use words, constructed into sentences and demarcated by punctuation marks. Our words are constructed from a 26- letter (or character) alphabet, and our number system, called the decimal system, is based on 10 digits, zero through nine. Through various combinations of these letters and digits, we are able to construct all of the words and numbers we use.

1.8 Computers, however, use a much simpler system, known as the binary system, to represent all of the data that they process. The only values in a computer system are 0 and 1. Thus all letters, numbers, punctuation marks, and so on are represented with unique combinations of 0s and 1s. In order for computers to manipulate the types of information with which we are familiar, the information must be converted to a form that the computer can understand.

1.9 The smallest possible unit of information in a computer system is called a BIT (BInary digIT). All that a BIT can signify is the difference between two opposites or the presence or absence of something. A BIT contains very little information which is similar to natural language where one letter contains very little information. However, BITs can be combined into unique patterns to represent letters, numbers, and symbols.

1.10 The combinations of BITs which represent numbers, letters, and special symbols are known as BYTES. There are three common codes for representing natural language in the form of bytes. The first is BDC (Binary Coded Decimal), a six-digit code, meaning six bits per byte. The second is EBCDIC (Extended Binary Coded Decimal Interchange Code), an eight-digit code. The third is ASCII (American Standard Code for Information Interchange).

1.11 Like the hierarchy of our natural language which builds words from groups of letters, sentences from words, paragraphs from sentences, and so on, computerized data are also constructed in a hierarchy. Bits are first combined to form bytes which are equivalent to characters in the alphabet or digits in the decimal number system. The next step in the hierarchy is the data element. This is a collection of bytes which refer to one item of information such as

a name, a social insurance number, or value for a certain entity. The physical space occupied by a data element is called a field. Indeed, the terms data element and field are often used interchangeably.

1.12 The next step in the hierarchy is the record or logical record. This is a group of related data elements that refer to one person, place, thing, or event. A group of logical records, which contain the same data elements in the same arrangement, is known as a file, often referred to as a machine-readable data file (MRDF). A file in a computer system is similar to a file folder which contains a group of identical forms. The last step in the hierarchy is the data base which is a collection of carefully integrated files, usually stored in a central location and accessible to a variety of users. (1)

**Storage Media**

1.13 Storage media are the materials on which data are written and stored. There are several common types: punched cards, magnetic tape, disks and drums, diskettes and cassettes. These storage media are sometimes called "data carriers" as they provide a way to carry data from input devices into the computer's main memory. Another function of storage media is to hold data in machine readable form outside of the computer's main memory. This reduces costs, as maintaining data in the main memory is expensive, and it also provides a method for security and preservation of the data.

1.14 A card reader is a simple mechanical device that can recognize holes punched in a computer card and transmit their meaning to the CPU. However, there are a number of drawbacks to the use of punched cards. Card readers are very slow compared to tape or disk. Furthermore, data to be entered must first be coded on cards by using a keypunch device, then transferred to the card reader. With this two-step system cards can be lost or shuffled, causing errors that can be very difficult to identify and correct. Cards are bulky and cumbersome to store and carry around, and must be treated with care to prevent damage.

1.15 They do have one unique advantage, however. The punched card is the only form of computer input that can be physically read and manipulated by humans. For this reason, cards will probably continue to be used for applications such as billing systems, where computer output must be distributed to many people, then returned to the computer. But for most applications the card reader is being rapidly replaced by data entry devices that allow data to be keyed directly to tape or disk, or by terminals that allow direct interaction with the computer.

1.16 The read/write head of a tape drive records data on iron oxide coated plastic tape as patterns of magnetized spots. Modern tape drives can read or write data at densities of 1600, 3200, or 6250 characters per inch.

1.17 At a density of 6250 characters per inch, a standard 2400 foot reel of magnetic tape could theoretically hold over 180 million characters. In actual fact, the capacity is somewhat less than this figure. A tape starts and stops frequently as it is processed. With each start, about five-eighth's of an inch of tape passes the read/write head before the device reaches the correct speed for reading or writing. Since data can be read or written only at the correct speed, these portions of tape, known as "inter record gaps", are not used.

1.18 In addition to large storage capacities, tape has the advantage of speed. A tape drive passes tape very quickly, transferring information to and from the CPU at very high speeds. Tapes are also inexpensive to purchase and compact to store or transport. For economical mass storage and fast retrieval, tape is the most commonly used storage medium.

1.19 There are, however, two disadvantages to the use of magnetic tape. First of all, tape can only be read sequentially. Because of this, small amounts of data cannot be retrieved without reading the entire file. Reading thousands of records on a tape in order to retrieve just a few is both a slow and an expensive means of retrieval. In addition, if information on two tapes is to be matched or compared during processing, the data on both tapes must be in the same sequence. Secondly, every time a tape is required an operator must first locate the tape, then mount it on the tape drive. For data accessed frequently, this is extremely slow and inconvenient.

1.20 Disk drives belong to a class of devices known as "direct access storage devices." Because they are faster and able to store more data than the other direct access storage devices, disks have almost entirely replaced drums. Disks are circular metal platters coated with a magnetic material. A number of disks (usually six or eleven) mounted on a central spindle constitute a disk pack. The surface of a disk is divided into a series of concentric circles, known as "tracks." Data are recorded on these tracks as patterns of magnetized spots.

1.21 A disk pack is mounted on the drive spindle, where the disks revolve at very high speed. The read/write heads are mounted on comb-like arms that are inserted between the disks. Each arm has two heads: one to use on the disk above it, and one to use on the disk below. The top surface of the top disk and the bottom surface of the bottom disk are not used. As the disks spin, data are passed by the heads where they may be read or written. The arms move back and forth to serve all the tracks.

1.22 In addition to the full size metal disk, there are also small plastic disks, known as "diskettes" or "floppy disks." The latter resemble small 45 rpm records sealed on protective paper envelopes. Diskettes are not combined into disk packs. Although their storage capacity is less than that of the larger disks, they are very inexpensive to purchase and are easy to use and store. Diskettes are used primarily in mini-computer systems, or as auxiliary storage for terminals or text processors.

1.23 With disk storage, the two main disadvantages of tapes are overcome. Although disk packs can be removed from their drives, many remain permanently mounted. The data on these permanently mounted disks can be accessed without the delay involved in locating and mounting a magnetic tape.

1.24 The second advantage of disk is the fact that data on disk may be retrieved by methods other than sequential processing. Access methods exist which allow records to be extracted without reading the entire file, and permit the insertion, change, or deletion of records without the need to make a new copy of the entire modified file. Furthermore, files may be read in several different sequences.

1.25 Although disk storage has advantages lacking in tape, disk packs and their drives are much more expensive than tape, and are also usually slower. Most large computer installations make use of both devices - storing most data on tape and using disk for small, frequently used files or files that cannot be processed sequentially.

**Hardware**

1.26 As mentioned previously, there are two kinds of hardware: the CPU and peripheral devices. They are used to provide the computer with input and to receive output from the computer.

1.27 The CPU is divided into two parts. One part, called the Arithmetic and Logic Unit (ALU), carries out all of the arithmetic and logic computations. The second part, the Control Unit, directs and coordinates the events taking place in the overall computer system. For example, it controls the moving of data from the input stage, through the program operations, to the output stage. The computer's Main Memory may be thought of as a large set of boxes,

like those in a post office. Each box has a number which is called an address. By using the address, the CPU can locate any box it wishes and use the data found in that box to carry out its processing operations. In addition to data, these boxes may also contain Instructions which tell the CPU how to process the data found in the other boxes. A group of instructions make up a Computer Program.

1.28 The Arithmetic and Logic Unit and the Control Unit of the Central Processing Unit consist of series of hardwired logic circuits. Each of these circuits is designed to perform a specific task and is incapable of doing any other. A circuit designed to add two numbers, for example, cannot be used to compare two numbers.

1.29 The remaining portion of the CPU, the Main Memory or primary storage, differs considerably in its requirements from the secondary storage media (such as tape and disk). Data recorded on tape or disk may remain on that tape or disk for minutes, days, or years. Data recorded in primary storage may only remain for a fraction of a second, as programs and data are only retained in primary storage while they are actually being processed.

1.30 To be efficient, primary storage requires a storage medium that allows data to be changed, tested, and moved at extremely high speeds. At the present time, there are two different types of primary storage in use. The first type is called core storage. Cores are tiny iron rings that can be magnetized in two directions. A computer's memory can be made up of millions of these cores, strung together on wires. Each core has a number of wires passing through its centre. There are two wires to apply the current to magnetize the core in the desired direction, one wire to sense the direction of magnetization, and one to prevent the sense wire from reversing the direction of magnetization.

1.31 Newer computers have replaced core storage with sophisticated integrated circuits. These tiny circuits are products of intensive research in the field of micro-electronics. Integrated circuits not only store data in much less space than core, but they are also extremely reliable and long-lasting, and are easily replaced when necessary. Ongoing research in this field continues to produce new circuits which are increasingly smaller, faster, and cheaper.

1.32 In processing data, a program (or set of instructions) and a set of data are fed into the computer's main memory. When the computer is ready to run the program, the control unit retrieves an instruction from the main memory and decodes it. This is called I - time for Instruction Time. The program will often ask for some input which the control unit sends to the ALU. The period when the actual operation is performed by the ALU is called E - time for Execution Time. This process is repeated until all of the instructions have been executed on all of the data. Modern computers execute between five and ten million of these instructions per second.

**Software**

1.33 It is important to emphasize that computers do not, and cannot, solve problems. They merely follow instructions given them by the programmer. The instructions are simple and rudimentary. They are built into the machine by manufacturers. Programming is nothing more than invoking these instructions in a sequence that will solve the problem.

1.34 Machine language is the lowest level of programming languages. It is, in fact, the language with which the machine operates. No matter in what language the program is written, it will eventually have to be translated into machine language. As computers are only capable of using the binary number system to represent all data and instructions, machine language is in binary. This means that codes to represent instructions and all addresses of storage are in binary. In the early days of computers this was the only language that could be used for programming. It was a very laborious method which was not satisfactory. The proposal was made to develop languages that were easier to write. The first development was the design of Symbolic Languages.

1.35 If a program could be written in a language that was easier for the programmer to write, it could be put through a translation program and translated into a machine language which the machine could understand (that is, binary). These symbolic languages became known as assembler languages and soon replaced the machine language as the lowest level at which a programmer would write a program. An important characteristic of symbolic or assembler languages is one instruction in these languages is translated to one instruction in machine language.

1.36 Although assembly language programming was easier than machine language programming, it was still a time-consuming and tedious task. In order to make programming easier and less time-consuming, high level or procedure oriented languages were developed. Today most programming is done in these higher level languages, and in the future less and less will be done in assembler. There are more than 100 programming languages in use today. The two most common programming languages are COBOL and FORTRAN.

1.37 COBOL (Common Business Oriented Language) was designed for business applications and is used throughout both the public and private sectors (that is, both by governments and private industry).

1.38 FORTRAN (Formula Translation) is the most popular language for scientific purposes. Programs written in FORTRAN are generally those that have little input and output, but carry out a great number of complex calculations during their execution.

1.39 When a program is written in any programming language, other than machine language, it is necessary to translate the program into machine language. When using high level languages (such as COBOL and FORTRAN), a special program called a "compiler" is used to carry out the translation process. In fact, a compiler does more than just translate instructions. It translates each data name or symbolic name to a machine code address and provides its length in bytes; it builds machine code labels for instructions that are "branched" during execution of the program; it translates each high level language instruction into one or more machine language instructions; and it provides diagnostic information to the programmer (that is, it checks to see if the programmer has followed all the rules of the language used and informs the programmer when it finds errors or suspects errors which will result in the machine language program malfunctioning).

1.40 Once a program has been compiled and tested, the machine language "object module" is placed on some medium such as tape or disk in a program library to be used as required without going through the translation process each time. When it is necessary to modify a program, the programmer changes the high level language program which is then compiled and tested and replaces the machine language program in the program library.

1.41 Today most users do not have to learn programming languages because they rely on software packages, also known as "canned programs." These packaged programs are stored in the computer's memory and they allow the user to use instructions that closely resemble English. Some, for example, permit a user to alphabetize or sort, simply by depressing a key. This then activates a stored program. Most packaged programs are based on a programming language and use a compiler to translate from the simple instructions to the programming language on which they are based.

## THE NATURE OF MACHINE-READABLE RECORDS

1.42 In managing information to date, the emphasis has been on the management of the physical entities on which the information is recorded. The information and the physical storage media which supported it were created once and then frozen in time and context. Changes, deletions, or additions which were made required another creation (or recreation) process involving another (additional or new) combination of information and physical storage media.

1.43 Even the introduction of microfilm and microfiche systems has not changed this basic concept. The term "updatable" microfilm or microfiche is really a misnomer. A fiche or film functions basically as a file folder in the traditional sense. Documents in the file (that is, photographic images) are not updated in the electronic data processing (EDP) sense. Images which are no longer needed or which are inaccurate are so identified (blocked out or the photographic image burned out), with the new image being added to the fiche or film. Physically, then, the old document continues to exist but no longer in readable or usable form, while new documents are added to the fiche or film, usually in sequential

fashion. Computer driven fiche or film retrieval systems are basically automated indexing/retrieval systems for locating the physical fiche or film (that is, the files not the information itself).

1.44 On the EDP side, it is the information itself as found on a physical medium which is updated. The physical medium itself remains as a specific storage location whose contents are changed (that is, the information is destroyed and/or replaced, not the physical storage medium). As such, technology has introduced a different type of information, namely "processable information." The information is accessible, interpretable, manipulable, and transmittable only by automated or electronic means. Information of this type does not exist as a defined and static set of data frozen in time on a specific physical medium, but should be viewed as a dynamic entity having certain organic properties being composed of unique, fundamental, and discrete bits of information or data elements which can be rearranged, changed, manipulated, merged, or deleted in order to generate a set of information on demand.

1.45 Processable information or machine-readable information is data created with the use of a computer which requires access to a computer in order to be transformed into a form that is readable by people. As we have seen, machine-readable information can be recorded on a variety of physical media-punched cards and magnetic media such as tapes, disks, drums, and diskettes. The contents of the information can range from the text of a letter, to detailed accounts of receipts and expenditures, to responses to survey questionnaires, to complex patterns of digits that represent the series of coordinates which constitute a map. (2)

## UNIQUE FEATURES OF MACHINE-READABLE RECORDS

1.46 Machine-readable records have characteristics which pose a variety of special problems for archivists. First of all, such records do not have the same kind of durability and lifespan as their hardcopy counterparts. It is not possible to leave magnetic tapes, on which most machine-readable records are stored, in dormant storage for a number of years before giving them adequate conservation treatment as is often the case with textual or other hardcopy cultural properties. Such maintenance measures as cleaning, precision rewinding, and copying prolong the life of the storage medium and thereby preserve the information on the tapes.

1.47 Machine-readable records can be easily updated, copied, erased, and reformatted. Some of these functions can be performed without leaving any evidence that changes actually occurred. It is imperative, therefore, that those who design computer systems be made aware of the value of some of the information, and that steps be taken when designing such systems to ensure that information of long-term value is properly identified and retained before it is transferred to the appropriate archival repository.

1.48 It is not sufficient to have access to the machine-readable information or raw data themselves. In addition to the data, it is necessary for an archivist to obtain documentation which describes the contents, arrangement, codes, and technical characteristics of the machine-readable data file. Without such documentation an archivist is unable to appraise the value of the information, and a researcher is unable to access the information.

1.49 The organizational environment in which machine-readable records are created and maintained is quite different than that in which hard copy materials are produced. Archivists who deal with machine-readable records must coordinate the activities of three groups - the creators of the records, the users, as well as the data processing personnel. When undertaking inventorying and scheduling activities, archivists must work with these groups, as well as with records management personnel. (3)

## SOURCES OF MACHINE-READABLE RECORDS

1.50 As we move towards an information society, the creation and use of information in electronic form becomes an increasingly significant activity in society. Governments are the largest single creators and consumers of information in most countries. Source data created at public expense may have applications in research and analysis, and form a substantial base for recording the development of a society. In the early days, applications of computer technology were related primarily to "number crunching" - that is, to business and financial systems as well as scientific analysis. Cultural effects were not readily apparent. Increasingly, however, various facets of the development of society are being recorded in electronic or machine-readable form.

1.51 The advent of automated or electronic data processing has led to the creation by government agencies of numerous data bases containing vast amounts of information about many aspects of a nation. There are census files, national and regional economic aspects, employment files and records of education, and statistics on crime, on the incidence of disease, and on the distribution of health resources. There are data bases on climate, weather, geology, on food production and food consumption, on transportation and communication, on the cost of living, and on the state of the art in many fields of science, medicine, and technology. There are files on the environmental impact of technology, on product safety, housing construction, the growth of forests and the decline of cities, and many more on the activities and expenditures of government itself. The list goes on and on, and it is still growing. As the ultimate repositories for all valuable records of national governments, national repositories, as well as those repositories of other jurisdictions, are in a crucial position to ensure that an adequate stock of information is preserved for the future.

## USES OF MACHINE-READABLE RECORDS

1.52 As one author has noted, "A measure of the relatively short time period between technical improvements in computers and their adaptation in varied fields is illustrated by how soon these expanded, were utilized for calculations, correlations and anlyses by economists, political scientists, sociologists, historians, jurists and linguists, in roughly that order." (4)

1.53 It was in the mid-1960s that political scientists developed techniques for analysing and recording numerous variables such as opinion polls and voting behaviour. As a result of this new technique for conducting political research, a number of academic institutions began to store opinion surveys in machine-readable form as well as to convert political data in traditional form to various tape formats. This resulted in the establishment of machine-readable archives in a university setting. This also led to the inclusion of computer programming in the curricula for advanced degrees in the social sciences. (5)

1.54 Sociologists, as well as historians, soon began to use the economic, political science, and other techniques for "recording, storing and tabulating quantitative data about individuals, institutions, events and goods to test earlier hypotheses and develop new insights." Jurists also discovered that the computer could be used in the search of laws and decisions on a wide variety of legal topics, and linguists, too, began to use the computer for analysing the use of words by different individuals. (6) Indeed, there are few, if any, disciplines which do not now use the computer in some aspect(s) of their research and analysis.

1.55 During the past two decades, graduate curricula in a large number of disciplines at many univesities have been modified by the addition of courses and seminars in quantitative methods. There have been numerous conferences held which in one way or another have dealt with quantitative studies. A great deal of time and attention has been devoted to the development and dissemination to scholars "of major collections of computer-readable historical research data and of specialized computer programs for analysis and manipulation of such data." (7) New associations and organizations have been formed which deal with quantitative analysis, computer systems, and various types or kinds of machine-readable data (such as numeric, cartographic, and textual). And there are a multiplicity of journals and newsletters produced each year which deal with various aspects of quantification.

## SUMMARY

1.56 The first part of this chapter has provided a brief overview of a computer system in terms of hardware, software, and data, an explanation of the input, processing, and output functions of a computer system, as well as a description of the various forms or

media on which machine-readable data can reside. The purpose of this section is to serve as background information and to provide clarification for terms and functions which will be discussed in later chapters, particularly chapters four and five.

1.57 The second part of the chapter has outlined a number of unique features or characteristics of machine-readable records to demonstrate the importance of developing new archival approaches and procedures and modifying existing ones. Other unique characteristics will be discussed in later chapters, but these will have a more direct bearing on the appraisal function itself. This section has also provided a brief outline of the large amount of information now produced in machine-readable form, as well as a short description of the various uses of such information.

## NOTES

1. For this portion of the chapter, the author has drawn heavily from a "Basic Computer Concepts Workshop" package prepared by Margaret L. Hedstrom of The State Historical Society of Wisconsin. The workshop was presented at the 45th annual meeting of the Society of American Archivists held in Berkeley, California, September 1981.

2. These ideas have been extracted from a report written by Jake V.Th. Knoppers, Managing the Electronic Revolution, Archiving the Electronic Heritage: A Position Paper on Issues and Problems in EDP Records/Data Management, prepared under contract for the Public Archives of Canada, Ottawa, December 1981, pages 7-8.

3. For a more detailed description of these and other characteristics unique to machine-readable records, see Archival Preservation of Machine Readable Records: The Final Report of the Wisconsin Survey of Machine-Readable Public Records. The State Historical Society of Wisconsin, Madison, 1981, pages 7-8.

4. Meyer H. Fishbein, Guidelines For Administering Machine-Readable Archives. Committee on Automation, International Council on Archives, Washington, D.C., November 1980, page 5.

5. Ibid., pages 5-6.

6. Ibid., page 6.

7. Jerome M. Clubb, "Quantification and the 'New' History: A Review Essay," The American Archivist, Volume 37, Number 1, January 1974, pages 16-17.

## Chapter II

## MANAGING MACHINE-READABLE RECORDS

2.1 Before outlining the procedures or steps which should be followed when appraising machine-readable records, it is necessary to discuss a number of factors which could have a major impact on an archivist's ability to perform this function properly. It is possible that some of these factors will not apply to all archival repositories, or they may affect some repositories in different ways. To a great extent, they reflect certain problems which archivists in the Public Archives of Canada have experienced over the past decade in their efforts to appraise machine-readable records created in the Canadian federal government. However, every effort has been made to provide examples from other countries where such information has been brought to the attention of the author.

### IDENTIFICATION OF MACHINE-READABLE RECORDS

#### Legislative Authorities

2.2 Many archival institutions are unable to deal with machine-readable records because of limitations imposed by statutory or other regulatory authorities. These limitations are basically twofold: restrictions on the type or kind of records which the repository can acquire, and restrictions on the accessioning of recent records.

2.3 With respect to the limitation on the kind of record which may be acquired, remedial action should be taken, if at all possible. Fortunately, rapid changes in the recording of information are automatically altering such limitations. For those institutions, however, which may still be in doubt as to their ability to acquire machine-readable records, it may be necessary to revise the definition of what may be acquired to include all information, regardless of the physical form or medium on which the information resides.

2.4 This particular solution is not only necessary to assist archivists, but also to convince the creators and custodians of machine-readable records that the latter are, indeed, public property. Perhaps because of the short-term use of machine-readable information, the fragility of the storage medium, the storage location of the information (in tape libraries or data processing centres physically removed from other records), and the ability to erase, update, copy, and reformat the data with great ease, many creators, users, and custodians of such information have an erroneous belief that the information is their own and not the property of the jurisdictional authority for which the employees work.

2.5 With respect to the limitation or restriction on the accessioning of information that may be 20, 30, or 40 years old, the solution may not be so straightforward. One thing is certain, however, and that is the fact that machine-readable records of long-term value cannot be left for such lengthy periods of time without certain precautionary measures being taken. Otherwise, the data might be erased, or they might be unreadable, or the written documentation which must accompany the data might no longer be extant.

2.6 In his Guidelines For Administering Machine-Readable Archives, Meyer Fishbein offers three suggestions to resolve this particular problem. The first is the "establishment of departmental tape libraries for the preservation of significant information in machine-readable form." The second is the "establishment of a central organization for such preservation." And the third is "a change in the enabling authority to permit the archives to accept a ... copy of any machine-readable [data] file that is appraised as having permanent value." (1) As Fishbein indicates, the third option is the best solution. Otherwise, there could be a split between traditional and non-traditional archives.

2.7 Another possible solution is the development of standards for the storage of any machine-readable records appraised as having permanent value to which creating departments and agencies must adhere before the records can be transferred to the archival repository. Such standards should address the need for related written documentation to be retained, as well as the need for the machine-readable data to be kept current with any hardware and software developments that might occur during the interim in the creating department or agency.

2.8 Assuming that the archival repository has the legislative mandate to acquire machine-readable records, it may be that this legislative authority is overridden by other government acts which might prevent archivists from both appraising such records or acquiring those which have permanent value. It is not uncommon in many countries to find acts governing national statistical agencies, revenue or income tax agencies, defence and police agencies, and so on, which may make it extremely difficult, if not impossible, for archivists to review the information created in such agencies or for permanently valuable information to be transferred to the archival repository. Although this particular problem is certainly not unique to machine-readable records, the situation is more exacerbated with this kind of information because so much of it is of a personal nature.

2.9 A number of solutions can be suggested to resolve the problem but, in the final analysis, it will depend upon the situation in each country. With respect to the ability of archivists to appraise machine-readable records in such agencies, it may be possible for archivists to be "sworn in" as temporary employees of the creating agency, thereby being subject to the same rules and regulations as the permanent employees of the agency. Should this not be possible, then efforts should be made to have the archival legislation take precedence over other acts and regulations with respect to access to records for appraisal purposes and the eventual transfer of the permanently valuable records to the archival repository. Or alternatively, steps could be taken to ensure that when certain acts are drafted or amended, the name of the archival repository appears where access to records, etc. is outlined. If these efforts are unsuccessful, then consideration might have to be given to the establishment of "mini machine-readable archives" in some of these agencies. However, the standards and procedures followed for the processing, description, and preservation of the records should be established by the archival repository, and the staff undertaking these tasks should be trained by professional archivists. Indeed, every attempt should be made to ensure that such employees are professional archivists and that they have some functional relationship to the archival repository.

Machine-Readable Data Created by Different Governmental Jurisdictions

2.10 Another problem which archivists appraising machine-readable records frequently encounter is that data held by a central government agency are often the aggregation or compilation of data created by other government jurisdictions. Three issues arise with respect to such data: the question of access, who "controls" the data, and the question of ownership. While this problem may arise more frequently in countries with federal forms of government, it can also arise in countries with unitary systems of government as well.

2.11 Certainly within federally-structured countries, there are many joint federal-provincial or federal-state programmes which result in the creation and even "sharing" of records. And yet the majority of the agreements which establish such programmes make little, if any, mention of the disposition, etc. of the records. As a result, the records can be subject to both federal and provincial/-state regulations. Conflicting records schedules and archival limitations at the federal and other levels of government can complicate the situation even more.

2.12 As for machine-readable records, it is quite possible to find copies of data files specific to each province or state in each of these jurisdictions as well as at the federal level, the federal department or agency having merged the various data files in order to undertake particular kinds of analyses, etc. Of particular concern to a national archival repository is the decision-making process which determines which of the data files are under the control or ownership of the federal government institution. It is this decision-making process which determines the limits as to which data files the national repository may acquire.

2.13 In order to resolve the problem, arrangements should be made whereby when such agreements are drawn up, specific mention should be given to the retention and disposition of any records which may result from the agreements. Such clauses should then be reviewed and approved by the appropriate archival authority. It may be appropriate for all existing, as well as new, agreements which result in the creation of machine-readable records to be reviewed in order to determine the precise status and ownership of such records. Archival limitations should then be established with the archives at other levels of government. For example, records identified as historically valuable and of national significance should remain in the national repository, while records of local or regional interest would be made available to the archives at other government levels. In cases where the records are of both national and local value, copies of the records could be made available to all of the archival repositories concerned. It is important to remember that machine-readable records can be easily and inexpensively copied, and that there is no such thing as an original machine-readable record, at least in the sense of a traditional manuscript collection. It is for this reason that the disposition of machine-readable records should not raise the jurisdictional issues which often occur with the more traditional hard copy archival materials.

2.14 There is an urgency in resolving the problem of federal-provincial or federal-state machine-readable data because of the rapid shift to on-line, real-time computer applications involving the sharing of machine-readable data for the management and delivery of shared (funding or jurisdictional) socio-economic programmes. A well thought-out implemented data base management system for such programmes would be of considerable assistance in establishing "control" or "ownership" of the machine-readable data.

## Machine-Readable Data Created as a Result of Government Contracts and Grants

2.15 Government departments and agencies commission hundreds, if not thousands, of very costly surveys or projects which are undertaken by the private sector, either firms or individuals. For many of these surveys and projects, data are collected and processed into machine-readable form. Often the data have archival value which is at least equal to, if not greater than, the report submitted which summarizes the data in aggregate form. In other words, much of the machine-readable data produced as a result of government contracts or grants form a resource which can be reused at marginal extra

cost. Such data may have a usefulness outside the particular survey or project for which they were created. In fact, the data collected or prepared could be of interest to other researchers or the public at large. The information could also be of archival value. It is therefore important that, where machine-readable records are created as a result of a contract or research grant, the records and any accompanying documentation, in addition to the final report, be delivered to the contracting government department or agency. Indeed, situations have occurred where a researcher wishing to use machine-readable records created by a contractor for a government agency has ended up paying the contractor for access to and a copy of the machine-readable records for whose creation public funds had already been paid.

2.16 In most cases, all data collection activities by the private sector, for which payment comes through some form of government funding, involve a contract of some sort. Most of these contracts should contain a clause stating that all records created as a result of the contracts become the property of the government institution which established the contract. The nature and possible variations in wording of these clauses could vary, depending upon the type or kind of contract. In any event, the records produced under contract should be received by the records manager of the responsible government department or agency for scheduling purposes. An archivist is then in a position to appraise the scheduled information.

2.17 The situation is equally confusing with respect to machine-readable records created as a result of government grants and similar government funding mechanisms.

2.18 In the United States, there is no overall policy as such with respect to machine-readable records created by such means, although some federal government agencies do require the deposit of the machine-readable records. (2) In England, the Social Science Research Council requests that the recipient of an SSRC grant deposit a copy of the machine-readable data files and supporting documentation with the Social Science Data Archive at the University of Essex, with terms of release and disclosure, favouring use of the data by other researchers as early as possible. Should the machine-readable data files in question not be deposited, this weighs very heavily in the evaluation of a further grant request by the applicant (that is, no deposit, no further grants). (3) In the Netherlands, the equivalent of the English Social Science Research Council requires the deposit of a copy of the machine-readable data files and supporting documentation with the Steinmetzarchief in Amsterdam.

2.19 In both England and the Netherlands, good working relationships appear to have been established between the granting bodies and the data archives. The latter often assist the researcher in data collection/preparation, design, and methodology, as well as in the preparation of proper documentation. The deposit policy also appears to be working fairly well, inasmuch as data are being reused, thus avoiding duplication. The granting agencies appear to be more willing to allocate funds involving the creation of machine-readable data files to those researchers with a proven track record.

2.20 In Canada, the two major federal government granting agencies in the communications/cultural sector are the Social Sciences and Humanities Research Council of Canada (SSHRCC) and the Canada Council. Both agencies are quite aware of possible impacts the use of computer- based technologies may have on the research process and the creation of machine-readable or digitized information of archival or cultural value.

2.21 The SSHRCC has just recently revised its "Research Grants - Guide for Applicants." The section on "Supplementary Guidelines" has been considerably revised and augmented to reflect current trends in research, research methods, and the use of computers in research. The "Guide" states quite explicitly that data collected in a survey supported by the SSHRCC are public property and not the property of the principal investigator. The data must be made publicly available within two years after the data collection phase is completed. The SSHRCC also expects researchers to deposit their data with an archives, data bank, survey centre, etc. where preservation and distribution can be ensured. The Council also requires applicants to indicate on their application the name of the organization selected for deposit and the terms of the agreement. It is also the responsibility of the principal investigators to prepare appropriate user and coding manuals. The Council is prepared, under certain conditions, to assume the cost of either making the data more usable to other scholars or preparing an anonymized data file.

2.22 The mandate of the Canada Council is to foster and promote the arts in the country. It has a policy whereby a copy of any machine-readable data file created as a result of a research contract with the Council itself must be deposited with the Council. Recently, the Canada Council established a section dealing with new technologies and integrated media. As such, it is now funding the creation of works of art in machine-readable or digitized forms. Obviously, these digitized art forms (that is, combinations of hardware, software, or data such as computer art, holograms, computerized animation, etc.) also form part of any country's "electronic cultural heritage." As such, it is necessary for archivists to become involved with such information.(4)

2.23 The general trend in research in the social sciences and humanities is to test one's hypothesis or findings on microdata, whether the microdata are observations on either present individuals (through survey questionnaires, etc.) or voluminous bodies of archival material containing microdata (such as parish registers, tax accounts, custom and port records, notarial records, etc.). In these areas the use of quantitative methods and the computer has almost become a sine qua non for social science researchers. The creation of machine-readable data very often involves a considerable utilization of resources, both human and financial. As such, a single MRDF can represent the results of tens of thousands of dollars of public funds and many person-years of work. The cost of making all these data files available to others can be very small in comparison. It is therefore necessary for archivists to ensure that steps are taken to identify and preserve those data files having permanent value.

## ESTABLISHMENT OF EDP RECORDS MANAGEMENT PROGRAMMES

2.24 "Processable information" or information in machine-readable form requires a different management approach than that for hard copy records. The approach of managing information through the expressions of its physical form is not a relevant approach for machine-readable records. While in traditional records management these have been sheets of paper, a file folder, a photograph, a map, an architectural drawing, etc., for machine-readable information the equivalents are, by and large, the data elements. It would therefore be appropriate to speak in terms of the development of a "data management programme" as the name for the policy guidelines which ensure that the basic principles of the management of recorded information are applied to machine-readable or processable information.

2.25 Regardless of the terminology which is used, it is crucial for EDP records management programmes to be established in order for archival repositories to be assured of having a systematic acquisition programme for machine-readable data. It is only in this way that archivists can properly identify and appraise the machine-readable records that are created in the particular jurisdiction in which they work. Just as procedures have been established for the identification, inventorying, and scheduling of hard copy material, so it is necessary to establish such procedures for the management of machine-readable data.

### Cost-Benefit Implications of an EDP Records Management Programme

2.26 One of the major rationales used for a records management programme has been the fact that certain savings could be achieved by storing voluminous quantities of records used infrequently by the user in low-cost storage sites (that is, records centres). It has also been demonstrated that intellectual organization of the information recorded on the physical storage medium could be classified and

arranged in a manner which would greatly facilitate retrieval. In most cases this has been achieved through the assignment of various physical identification attributes to the recorded information by placing it in file folders, on reels of microfilm, and so on, using colour coding and physical means of organization to manage the information.

2.27 Subject classification schemes, central and satellite records offices, filing and indexing schemes have been developed over the years. Consequently, a number of cost-benefit approaches have been developed to the question of central versus decentralized or no records offices, microfilming proposals, mail room operations, different subject classification systems, and so on.

2.28 Cost-benefit analyses for EDP records management, on the other hand, are still in the infancy stage. Analysing cost-benefit factors for machine-readable information in terms of paper equivalents can be a useful exercise if only to demonstrate the "density factor" associated with the machine-readable information. If one calculates in terms of file drawers and linear feet of storage space, it has been projected that the paper equivalent of one 2400 foot reel of magnetic tape, stored at 6250 BPI, is approximately 503 feet (or at 1600 BPI, it would be 125 feet). (5) One way to estimate the value of a single reel of magnetic tape would be to calculate the amount of resources (materiel, financial, and personnel) which would, on the average, generate or be supported by 503 (or 125) feet of textual records.

2.29 It is important to keep in mind that it is not the magnetic tape itself, but rather the information on the magnetic tape, that has the value. In fact, it can be easily argued that the cost of the physical storage medium for machine-readable data is a relatively insignificant cost factor in undertaking a cost-benefit analysis of an EDP records management programme. The costs involved in generating the data on the tape are of a marked higher order of values than the cost of purchasing the tape or storing it.

2.30 Another factor which should be taken into consideration is the amount of resources which a government or company spends on EDP. In many jurisdictions, policies are in place relating to EDP as a production facility. However, it is equally important to be able to ensure the efficient and effective use of this expensive production facility in terms of what data are created, processed, produced, and utilized. In other words, a fully operative EDP records management programme is mandatory.

2.31 There are indications in many jurisdictions that the annual growth rate in EDP expenditures is stabilizing. By approaching EDP expenditures from the monetary and personnel side only, with some consideration for hardware and software, the marked increased capacities in data processing and data storage, due to a technological revolution, remain hidden. For example, from 1975/

1976 to 1978/1979 inclusive, the industry average cost for electronic data storage decreased by a factor of six to one for magnetic tapes and by a factor of three to one for magnetic disks. (6) Therefore, for the same amount of 1975 dollars, one could, in 1979, store six times as much data on tape and three times as much on disks. Or stated another way, given a slight increase in monetary expenditure, but including the factor of technological advances, the storage capacity of machine-readable data doubled in three years. There is every indication that the size of an organization's "electronic filing cabinet" will continue to double every two or three years.

2.32 Few organizations have figures available, or indeed have established the procedures to obtain the figures, on the volume of machine-readable data which are "disposed" of on an annual basis. In actual fact, this particular approach is not that relevant to machine-readable records, for three basic reasons.

2.33 First of all, the storage medium is reusable. While the recycling of paper involves the storage medium physically leaving the system, to be resold as recycled blank paper at some time in the future, magnetic tapes do not need to leave the system.

2.34 Secondly, the "disposal" of machine-readable information is a process which takes place within the system itself. In other words, the information is deleted or disposed of at the data element level or entire files are purged as part of the updating process.

2.35 And finally, as discussed above, the quantity of machine-readable information stored on electronic storage media is increasing. With a magnetic tape, it is possible to go from 800 to 1600 to 6250 BPI, and possibly to larger block sizes in the future, thereby placing substantially more information on the same physical storage media. This means that the same number of magnetic tapes can now hold eight times as much information as they did fifteen years ago.

2.36 Only when magnetic tapes are worn out or disk packs become obsolete could they be considerd "disposed" of in the traditional sense.(7)

## SCHEDULING OF MACHINE-READABLE RECORDS

2.37 While disposal plays a major part in traditional records management policy and procedures (it takes time and effort to get rid of the material), it constitutes a very minor exercise in the EDP world. Left on their own, those who control computer systems would automatically delete unwanted or unnecessary information.

2.38 It is therefore important that records schedules for machine-readable information be established at the system design or planning stage for new applications or programmes. It is at this time that the reasons must be given or the arguments made as to why the information is to be collected or created. How and where the data are to be used will also be determined. The question also arises as to how long the data are to be used for administrative or operational purposes.

2.39 With the availability of such information, an archivist should be able to undertake a preliminary appraisal of the data in question. The tagging of certain data for permanent retention would then be built into the design of the system. The authorized disposal or destruction of data according to an approved records schedule, and the spinning off of data marked for permanent retention according to the archival limitation in the records schedule, would take place as part of the regular operation of the computer system. In effect, this would be moving towards automated records scheduling and archival acquisition.

2.40 It is important to note that when arranging schedules and retention periods for machine-readable data, one should use "cycles," the time period being determined by the computer system which supports a programme or a delivery of a service (such as unemployment insurance, pensions, and other socio-economic programmes) requiring personal information. Therefore, instead of "two years," the retention period would read "two cycles" (or two months for change of address, marital status, number of dependents, etc. where there is a monthly cycle) or "three cycles" (that is, six weeks for change of address, marital status, number of dependants, etc. where there is a biweekly cycle).

2.41 As has been mentioned previously, because of the fragile and "time sensitive" nature of electronically recorded data, archivists do not have the luxury of time as do their colleagues who deal with hard copy records. If machine-readable information is scheduled for permanent acquisition, it should be done at a time which would allow for data verification by means of obtaining copies or samples of transaction and audit files, both of which are generated by computer systems. In other words, there may be situations arise (particularly with respect to personal information in a dynamic and interactive computer system) where an archivist requires not only an historical or archival master data file to be produced, but also the creation or copying of certain transaction files in order to document what has occurred within the system. In most cases, such transaction or audit files would be sampled.

2.42 When discussing the scheduling of machine-readable records, it is important to draw attention to General Records Schedule 20 adopted by the National Archives and Records Service in the United States. (8) Schedule 20 is designed to serve as an equivalent to a "general

records disposal schedule" for machine-readable records. It covers the retention and disposal of machine-readable data files according to their subject content within a given system (that is, edit, transaction, master files, etc.). It also addresses the American federal government's property interest in machine-readable records generated as a result of grants, contracts, contributions, sales, or other federal and non-federal agreements. Some general and specific comments concerning General Records Schedule 20 are outlined in the next eight paragraphs.

2.43 Although four factors are identified in Schedule 20 which distinguish machine-readable records from records of other media, a fifth factor should also be addressed. It is important to consider the distinction between the retention and disposition of the magnetic media and the retention and disposition of the data on the media.

2.44 Although the definition of master file used in Schedule 20 is applicable to many single purpose computer applications, in the more complex data base environment it may be difficult to achieve. For example, inputs, transactions, and edits could be directly fed into output reports or tables, thus making the identification of master files difficult. Furthermore, in a data base environment, any number of logical master files could be defined from one physical data base, depending upon the structure of the data base.

2.45 On page two of the Schedule, the following statement appears: "The proper scheduling of processing files can increase the availability of space on machine-readable media and reduce agency expenditures for stocks of magnetic media." Needless to say, this should not be the only reason provided for a scheduling system. Emphasis should be given to the sanctioning of legally established retention periods, response to privacy issues (that is, agencies should not keep personal information for periods of time which are longer than required), archival review issues, and so on. The argument could also be presented that the sanctioning of retention periods by a central agency (such as the National Archives and Records Service) means that any accusations by the public concerning any supposed inadvertent destruction of data by a government agency could be deflected to the National Archives if the data have been scheduled and the schedules endorsed by the National Archives.

2.46 On a related point, it is doubtful if the National Archives and Records Service could realistically expect departments to establish retention periods which are dramatically different from the rentention periods or life cycles set out in the specifications of individual systems. Indeed, by setting standards for the retention periods of various processing files (such as "dispose of after three or more life cycles" or "dispose of when no longer needed"),

NARS may be coming dangerously close to setting standards for the development of government - based information systems, a situation which might not be easily accepted by the EDP community.

2.47 Many of the master files described in General Records Schedule 20 are, in fact, processing files. These include technical reformat files, print files, publication files, and sample and/or sub-sample files. The other master files are related, not to function or process, but rather to the subject content of the data. In the author's view, the mixing of retention provisions for the retention and disposal of both the subject content and the function or process of machine-readable data limits the value of Schedule 20. The subject content of the data should not be a factor in the determination of retention periods, except as it relates to various legal issues. Subject content descriptions should be attached to descriptions of system overviews in order to permit archival assessment, but should not be the determinant in setting retention periods. The functions or processes which carry the data from their first creation, to their use, and to their final disposition should be identified, and the length of time that each product of a specific process is held should be determined and recorded. Undoubtedly the retention periods and generic descriptions of the subject content of year-end master files will have much greater interest for archivists than will the products of edits and validations. In summary, the setting of retention periods should be related to functions or processes (as well as legal requirements), while the setting of archival limitations should be related to the subject content of the data flowing through the system.

2.48 A considerable number of processing files are described in Schedule 20. Associated with each file type are a set of retention and disposal conditions (for example, "submit an SF 115", "dispose of after 3 cycles", etc.). This approach seeks to emphasize the importance of "processing" files as potential archival records. It also attempts to establish a consistent set of retention and disposal practices for all processing files in all government agencies. As already indicated, this is both an ambitious and somewhat dangerous endeavour. Such an approach would conflict with those systems which, for example, cannot maintain valid transaction files for three or more update cycles. Instead of applying a set of retention standards to these files, the scheduling process should be directly related to the existing retention and disposal practices which are dictated by the system specifications. As mentioned previously, the scheduling of systems should be completed at the system design stage when both systems managers and records managers could review and sanction (and adjust where necessary) the retention periods for various stages or processes within a given system.

2.49 There is also a concern on the part of the author that the strong emphasis placed on the scheduling of processing files could detract from what should be viewed as the most important scheduling function - the establishment of retention and disposal specifications, as well as archival limitations, on the most important stages of the system. In many respects the scheduling of processing files is analogous to records managers attempting to establish schedules for the hard copy working files, drafts, and other "process" information which serves only as input to a "final" product (such as a policy, a memorandum, etc.). Many records managers would argue that they have enough problems attempting to schedule the so-called "final" product without having to worry about scheduling the individual "processes" leading to the final product. Similarly, it would be difficult to expect EDP personnel to respond to retention requirements for "processing" data files when they will have enough difficulties attempting to schedule "master" files.

2.50 Schedule 20 provides a most useful outline of the various forms of documentation which support a given information system. The categorization of this documentation is useful because it gives a clear understanding to EDP personnel of the extent of the documentation which must be retained and scheduled with a given system. This is extremely important with respect to the archival limitations which must be applied to specific data files. In order for the archival management system to function (that is, for the data to be preserved), the supporting documentation must be held with the data. This linkage can be achieved through the consistent scheduling of both the data and the documentation. (9)

2.51 A general disposal schedule for machine-readable data should be based on the life cycles of data flowing through systems. Such a schedule can only be successfully applied if it is not restricted solely to machine-readable data. Any EDP system will be supported by input and output hard copy records, as well as supporting documentation. Therefore, the input hard copy questionnaires, forms, and other data collection documents, as well as the output reports, listings, etc. should be scheduled in the same context as the machine-readable data which are created, retained, and disposed of as a series of steps within the system. This approach ensures that all of the intermediary steps which form the life cycle of the information flowing through a system are identified.

2.52 And finally, it must be stressed that the archival limitations for information in machine-readable form may often be different from those for paper records. Frequently paper records which fall into the "housekeeping category" are appraised as having no archival value, basically because the cost (time and money) of acquisition, storage, and retrieval (that is, research) involved are too great in comparison to the "useful information" that can be obtained. However, in machine-readable form the same information of low archival value is compacted in a dense and processable form which greatly enhances its value. Therefore, a general records disposal schedule for machine-readable records, while possibly having the same retention period regardless of physical storage medium, might have quite a different archival limitation from its paper equivalent. This is another reason why serious consideration should be given to the development of separate general disposal schedules for machine-readable information.

## Roles of Records Management and EDP Personnel in the Scheduling of Machine-Readable Records

2.53 At the present time the majority of records managers and records officers have neither the training, background, nor understanding of what machine-readable information as well as computer systems and processes are all about. There is no blame to be attached; it is simply a statement of fact. Similarly, the number of EDP personnel who understand, or are even interested, in records management principles and procedures is equally scarce.

2.54 Until such time as this situation changes, if indeed it does, the role of records managers in the scheduling of machine-readable information will probably remain minimal. For the time being at least records managers should concentrate on serving as the link to the archival repository and as the officers responsible for coordinating all formal requests for approval of records schedules, including those covering machine-readable information.

2.55 Traditionally, EDP managers have concerned themselves with the establishment and management of EDP production and processing facilities. They have been involved with hardware and software acquisition, rationalization and harmonization of the varying needs of EDP users, ensuring that the integrity of the computer systems is maintained, and that communication networks and other data exchange protocols function properly. They have shown little inclination over the past twenty years (in fact, have had their

fingers burned) if they became involved in any noticeable degree with the contents of the information processed in the computer systems. Indeed, it has been the users of the EDP facility who have traditionally specified what information is to be processed and stored and who was responsible for tracking it (that is, which tapes and data files were his/hers and whether the data should be purged or kept).

2.56 However, at least in the Canadian government, there are indications that these attitudes on the part of EDP managers are slowly changing. Some EDP managers are now demonstrating far more concern with data use and retention, as well as a greater inclination to police users of their facilities as to the activity or dormancy of their data files.(10) This is a most appropriate change, inasmuch as the EDP manager is the officer who usually has physical control of the machine-readable information.

2.57 It is therefore the EDP managers who should be made responsible for ensuring that all the machine-readable records in their institutions are scheduled. It is with these officers that archivists must work on a regular basis in order to ensure that machine- readable records are properly identified, inventoried, and scheduled, and that those records appraised as having permanent value are indeed transferred to the archival repository at the end of the retention periods and in the format(s) specified.

2.58 It must be emphasized, however, that it is the users of the EDP facilities who must determine what data are to be created, collected, processed, and stored. It is also they who must specify and justify how long data are required for programme delivery or analysis purposes.

**SUMMARY**

2.59 In the first part of the chapter, a number of issues have been outlined which could affect an archivist's ability to appraise machine-readable records. The importance of the archival repository having a mandate to accession machine-readable records has been emphasized, as have the problems which can arise when archival legislation is superceded or overridden by other acts and regulations. Examples have been cited of machine-readable records being created by different governmental jurisdictions as a result of formal agreements without any clarification as to the ownership of the records. The value of machine-readable records which are created as

a result of government contracts and grants has also been explained. In all of these cases, suggestions as to how the various problems might be resolved have been given.

2.60 The second part of the chapter has described the importance of an EDP records management programme in ensuring the establishment of a systematic acquisition programme for machine-readable records. Some cost-benefit implications of an EDP records management programme have also been outlined. Suggestions have been made as to how one might go about scheduling machine-readable records, with a critical analysis provided of General Records Schedule 20 used in the National Archives and Records Service of the United States. And finally, the roles of records managers, EDP managers, and EDP users in the scheduling of such information have been described, and the importance of archivists working closely with personnel who design computer systems in order to ensure that retention periods and any archival limitations are included in the systems specifications and the software programmes.

## NOTES

1. Meyer H. Fishbein, Guidelines For Administering Machine-Readable Archives, page 9.

2. Thomas E. Brown, "Who Owns Contract and Grant Data and Who Can Use It? A Look at the U.S.A.," IASSIST Newsletter, Volume 6, Number 3, Summer 1982, pages 4-8.

3. Cally Brown and Marcia Taylor, "Who Owns Contract and Grant Data in the U.K. and Who Can Use It?", IASSIST Newsletter, Volume 6, Number 3, Summer 1982, pages 9-15.

4. See Jake V. Th. Knoppers, Towards a Canadian Electronic Cultural Heritage/Vers un patrimoine Canadien informatisé, a discussion paper prepared under contract for the Department of Communications/Public Archives of Canada, Ottawa, May 1983, pages 8-10.

5. The Computer Systems Services Division of the Public Archives of Canada developed the framework, etc. for calculating EDP - paper equivalents.

6. Information provided by the EDP Centre of Statistics Canada. It is also interesting to note that Datamation reports a shift from 70% CPU and 30% memory of five years ago to 40% CPU and 60% memory by 1983.

7. Much of the information concerning cost-benefit implications has been drawn from Jake V. Th. Knoppers, Managing the Electronic Revolution, Archiving the Electronic Heritage: A Position Paper on Issues and Problems in EDP Records/Data Management, pages 10-14, 50-51.

8. General Services Administration, National Archives and Records Service, General Records Schedule 20, Machine-Readable Records, FPMR 101-11.4, February 16, 1977.

9. The author is indebted to John McDonald, the Chief of the newly established EDP Information Systems Section of the Pulic Archives of Canada, for this analysis of General Records Schedule 20. Part of the responsibilities of this Section over the next year will be the development of a General Records Disposal Schedule for Machine-Readable Data.

10. This interest is best demonstrated in the increased use of "data dictionaries" throughout federal government departments and agencies. A data base usually consists of three components: a data dictionary

which defines the structure of the information; a data manipulation language which enables data to be entered, modified, or deleted; and a query language which enables a user to retrieve information from the data base. Usually a data base also has a communication capability. A data dictionary defines every element of data that is within the data base giving field length, format, field usage, and who may have access to the data. It is also used to establish logical groupings of data elements and to establish linkages between these groupings. Apart from its use internally within the computer, it will also serve as an external document showing who is responsible for the generation of the data, and how errors in source data are to be corrected. It is a device best suited to eliminate data redundancy and to control the use of standard data definitions. A data dictionary serves as the central control function which will determine data standards and enforce them. Only such a centralized control function will ensure the integrity of the information contained in the data base. The data dictionary could also contain a records scheduling (how long to be kept) and archival component (what should happen to data about to be deleted).

## Chapter III

## APPRAISAL OF MACHINE-READABLE RECORDS: PART I

3.1 Appraisal is the "process of determining the value and thus the disposition of the records based on their current administrative, legal, and fiscal use; their evidential and informational or research value; their arrangement, and their relationship to other records."(1) The appraisal of records is the most difficult and certainly the most important activity undertaken by an archivist. Future research is, to a great extent, dependent upon the judgments of today's archivists. It is therefore crucial that this judgment be based on a thorough knowledge of the records, the workings and organization of the department, agency, or institution, the programme which the records supported, the information contained in the records, and how the records might be used by researchers. As Theodore R. Schellenberg has stated, "... Appraisals of records should not be based on intuition or arbitrary suppositions of value; they should be based instead on thorough analyses of the documentation bearing on the matter to which the records pertain. Analysis is the essence of archival approval."(2)

3.2 The appraisal of machine-readable records involves the evaluation of the information contained in the records (content analysis), as well as an evaluation of the technical aspects of the records (technical analysis). The content analysis involves the traditional activities of archival appraisal combined with some new considerations particular to machine-readable records. The technical analysis is a relatively new activity in the appraisal of records, but one which is of the utmost importance in the evaluation of machine-readable records.

### CONTENT ANALYSIS - AN OVERVIEW

3.3 An archivist who appraises machine-readable records must ask the same kinds of questions in determining the informational value of the records as does his/her colleague appraising textual records. Does the record have legal, evidential, or informational value? Does the record have unquestioned permanent value? Does the record have immediate or long-term research value?

3.4 Records which have evidential value are those which provide "evidence" or testimony of the existence of a department, its organization, functions, and activities. In order to judge the evidential value of machine-readable records, the archivist must have an in-depth knowledge of the internal organization of the department or agency, the programme it supports, and the activities it carries out. As one senior archivist has stated, "At the outset it

is important to emphasize that appraisals of evidential values should be made on the basis of a knowledge of the entire documentation of an agency; they should not be made on a piecemeal basis. The archivist must know the significance of particular groups of records produced at various levels of organization in relation to major programs or functions."(3) Machine-readable records may have evidential value if they contribute to the policies or decisions adopted by a department or agency, or if they provide documentation of significant operations or procedures. Examples of machine-readable records which may have evidential value are provided in the second part of this chapter which deals with the application of content analysis to individual categories of information.

3.5 Although there are few countries at the present time in which machine-readable records are admissible as evidence in a court of law, there are indications that this is changing. According to Meyer Fishbein, when archivists accession machine-readable and other nontextual records, they must consider the question of the admissibility of the records as evidence in court. With respect to machine-readable records, it is the support documentation which becomes mandatory, both to understand and access the data and to ensure their admissibility in court. As Fishbein has explained, "Records custodians must be prepared to explain to a court how data were entered, how they were processed, who had access to a system, and what kinds of information the system generated." Even "logs of users of a system" may become an important issue, as they contain the names of individuals who may have gained access to the computer and adjusted the software.(4) It is obvious that this level of detail far exceeds the kind of documentation which is normally required to appraise, process, and use machine-readable records, as shall be explained in the next chapter on technical analysis. Therefore, for information in machine-readable form which may have legal value, the appraisal process may require a page-by-page review of the documentation, as well as a review of the automated administrative records which monitor computer usage.

3.6 Another aspect associated with the legal value of machine-readable records is their connection with national, and eventually international copyright law. Although copyright legislation in many countries does not yet address issues associated with machine-readable records, in those countries with more up-to-date copyright ligislation, the copyright laws make special provisions to cover computer programs. Many institutions purchase or lease "off the shelf", copyrighted software. While most agreements with software vendors may provide the institution with the right to use and copy the software, there may be occasions when this does not occur. Should the repository require a copy of the software to access certain machine-readable data, it may be discovered that the original agreement forbids the duplication of the proprietary

software. Under such circumtances, it might become necessary for the creating institution or the repository to purchase additional rights to use the software package.(5)

3.7 There is an additional factor which must be considered when assessing the legal value of machine-readable records. In every country there are various national as well as provincial/state acts which stipulate that certain kinds of records must be retained for certain periods of time to meet particular legal requirements. These cover such areas as banking, education, health, welfare, trade, commerce, and so on. In most cases the acts have been written with textual records in mind, although some address the retention of microform records. However, as such acts are revised and updated, they are including in the definition of "records" which must be retained for certain periods of time EDP or machine-readable data with the supporting documentation. It is therefore most important that archivists be familiar with the various acts which are relevant to their mandated jurisdiction and, if machine-readable records are included in the various acts, then the archivists must consider the implications when assessing such records for their legal value. Such an assessment should be based on all the information which is produced, and a decision rendered as to which form of information (that is hardcopy, microfilm, or machine-readable) it would be advisable to conserve in meeting the stipulated legal requirements.

3.8 The main appraisal judgment in terms of content analysis is the value of the information the records contain for uses other than those for which they were created. The determination of informational value of machine-readable records is similar to the evaluation of other types of information for potential research value. In other words, it is necessary for archivists to evaluate the significance of the subject content for current and future research. Because the subject content may be in areas or fields with which the archivists may be unfamiliar, they may wish to consult with researchers who are more familiar with the subject area before rendering a final decision on the informational value of the records.

3.9 Several general points should be considered in the appraisal of machine-readable records for their informational value: the uniqueness of the information or its format; the importance of the information; the manipulability of the information; the level of aggregation; and the linkage potential. The uniqueness of the information relates to whether or not the information can be found elsewhere and in what form (machine-readable, textual, microfilm). If the records are only in machine-readable form, it is important to know if duplicate copies exist or if these records are an extract from a larger file. For example, information in many machine-readable data files is duplicated in at least one other file

and probably many other files. First of all, textual input documents will undoubtedly contain similar microlevel information, and printouts and reports will contain both microlevel and aggregated information. Secondly, most computer systems will consist of input files, transaction files, master files, extract files, summary files, and so on. Each of these files contains information which duplicates, to some extent at least, the information in the other files. And thirdly, because of the transportability of machine-readable data by means of telecommunication lines or magnetic tape, many computer systems "interface" with one another. For this reason, it is not at all uncommon to find information in one system that may be duplicated in part in another system in a different agency with a different computer application.

3.10 The importance of the information being evaluated is the most subjective aspect to be considered in determining the informational value of the records. It is here where experience, training, wisdom, as well as historical and archival judgment are crucial. It is also at this point that the archivist's knowledge of current research trends and methods and the research community itself will be a determining factor in the evaluation of the information. Knowledge about the use(s) to which the information was put (primary use) and the use(s) to which the information can be put (secondary use) are essential.

3.11 There have been occasions when textual records have not been acquired because they were considered to be too cumbersome to use profitably for research purposes and they were also far too great in extent. However, the same information in machine-readable form can be manipulated easily. The information can be quickly retrieved by any variable or combination of variables and analysed by a variety of mathematical and statistical techniques. As a result, information which previously might have been assessed as having little archival value, due to the effort to use it, must now be examined in another light.

3.12 The level of aggregation is another important aspect to consider. Machine-readable records can contain microdata or non-aggregated data (that is, information on individual persons, things, and activities), or the information can already be summarized or aggregated. Machine-readable records containing information in a non-aggregated form have more potential research value than data which are aggregated. It is also true that a master file will have more research potential than an extract of the same file. When appraising machine-readable records, archivists should consider whether microlevel data, or aggregated data in the form of summaries and statistical reports, or possibly both, should be preserved.(6)

3.13 A further significant factor in the appraisal of machine-readable records is the potential for record linkage. This is when common identifiers such as name, address, social insurance or security number, and date of birth or common attributes such as sex, race, and age, are combined from two or more sources. Although record linkage is possible with texual records, it is a very time-consuming process. However, with machine-readable records the computer can be programmed to match cases which share common identifiers or attributes and, as a result, the linkage of thousands and even millions of cases becomes possible.

## CONTENT ANALYSIS - APPLIED TO INDIVIDUAL CATEGORIES OF INFORMATION

3.14 This section has been designed to assist in the more detailed analysis of informational and evidential value as they relate to individual categories of information. The categories have been developed in a somewhat arbitrary fashion, according to the subject classes of information most commonly managed in machine-readable form. Each category is further sub-divided according to the relative research values of various types of information associated with each. These values can range from "high archival value" to "extremely limited archival value." A "no archival value" is not acceptable.

3.15 Content analysis should be performed in consultation with departmental users, data processors, and other individuals connected with the information described in the file. This section of the chapter is intended as a guideline only and cannot be used without the archivist having first obtained detailed information on the organization, the information structure of the organization, the purpose of the file, the methodology used, its use in the specific programme, its relationship to other programmes in the organization, and even its value in terms of the user's own perception of its worth to both the organization and potential research communities.

### Administrative Data

3.16 Administrative data describe information collected and used by a department or agency to support the infrastructure within which the operational mandate of the department or agency is performed. These data are often referred to as housekeeping data.

3.17 The categories described in the following pages have been defined according to the types of data encountered by archivists in the Machine Readable Archives Division of the Public Archives of Canada. In this context they are only used to organize more clearly the narrative which follows. They are not designed to replace

other definitions. They are as follows: personnel; supply; financial; and project management.

**Personnel Data**

3.18 These are data collected to support the administrative requirements of a department or agency. They describe various attributes associated with the individual employees in the department or agency. The record lengths are usually long and contain such variables as name, home address, social insurance or security number, group and level, title, location, sex, branch and unit, date of employment, leave information, educational background, employment history, and so on. The time span of the information is usually dependent upon the length of time the individual is employed by the department. Although the number of records will vary according to the number of employees in the organization, most personnel operations are centralized and the systems, therefore, tend to be large. In most cases the format and nature of the variables of all personnel systems will be similar. Because of certain administrative requirements, departments may be required to maintain records for a specific period of time beyond their operational life.

3.19 These data may have evidential value because they are among the source records used to formulate human resources policy within a given department. They may also have informational value if they document a significant social group, the study of which could have significance for research concerns outside of those associated with the department. In many cases they may be expected to have some degree of potential for research beyond those associated with the administration of the department or agency.

3.20 The extent to which the data will be used will be determined by (in order of importance):

the significance of the attributes of each employee;

the number of variables of information provided; and

the uniqueness of the organization involved.

3.21 Their potential research value will be enhanced if they can be linked to other records. It should be expected that, although these records will be used primarily for statistical analysis, they may also be requested for individual record retrieval purposes.

3.22 **Examples**

3.22.1 Personnel data having high potential for research use normally describe, in considerable detail, the attributes of a socially significant group of individuals (that is, a defined group of individuals considered significant outside of their own administrative importance in the organization).

- military personnel system
- parliamentary personnel system

3.22.2 Personnel data having less potential for research use include data extracts from larger personnel systems describing a socially significant group.

- all senior managers in a department
- diplomatic service personnel from a foreign affairs personnel system

3.22.3 Personnel data having limited potential for research use usually support a limited number of variables describing individuals representing a wide range of occupational groups, educational backgrounds, employment histories, and so on. This is particularly the case if the system as a whole is similar to other personnel systems supported by a government.

- departmental personnel system

**Supply Data**

3.23 These are data collected to support the administrative requirements of a department or agency. They describe various attributes associated with the materiel used by a department to perform its operational mandate. The record lengths may vary but generally they are brief and may simply contain such variables as name of item, stock number, supplier, cost, number in stock, number on order, etc. The records tend to be repetitive and voluminous. Data are changed or updated frequently and a year-end master or other unique version of the data is rarely created. The time span of the individual records is dependent upon the length of time the individual supply items are used by the department. Since the approach taken to the management of supply data is common throughout a government or institution, most supply systems will employ similar record structures, variable definitions, and software techniques.

3.24 These data may have limited evidential value because they may form the source records used to formulate supply and materiel policy within a department or agency. They may also have limited informational value if they document a significant group of items, the study of which could have significance for research concerns outside of those associated with the department. In most cases, however, they will have little potential for research use beyond those associated with the administration of the department or agency.

3.25 The extent to which the data will be used will be determined by (in order of importance):

the significance of the supply item for research concerns other than those associated with the administration of the department or agency;

the relationship of the supply items to the operational policies of the department or agency; and

the uniqueness of the supply item to the organization involved.

3.26 Although most supply data will have limited research value, their value could be enhanced if it were possible for them to be linked to other records. These records will be used essentially for statistical analysis and rarely for record retrieval purposes.

3.27 **Examples**

3.27.1 Supply data having some potential for research use normally describe, in considerable detail, the attributes of a significant group of items (that is, a defined set of items considered significant outside of their own administrative value in the organization), which can be closely linked to the operational activities of the department or agency.

- National Defence radar equipment supply system
- Atomic Energy radio isotope supply system

3.27.2 Supply data having less potential for research use include data extracts from larger supply systems describing a significant group of items.

- National Defence supply system: heavy equipment tracked
- Department of Communications supply system: satellite communications equipment supply file

3.27.3 Supply data having limited potential for research use normally describe items which are more closely associated with the administrative procedures of a department. They also support a limited number of variables which in themselves are common to most departments and agencies.

- departmental office equipment supply system
- Post Office: fleet of delivery trucks supply system

**Financial Data**

3.28 These data describe information concerning the financial history, status, and future of an organization. They usually focus on such budgetary aspects as revenues, sources of funding, organizational structure, expenditures, overhead, commitments, debits, and so on. The record lengths are usually long and often contain considerable information on the functional attributes of individual programmes, various monetary details concerning these attributes and pro-

grammes, and financial projections covering the status of the programmes over a period of time.

3.29 The data are collected on a regular basis, usually according to a standardized format and usually on a frequent basis according to the financial activity (for example, accounts receivable, expenditures, and so on) experienced by individual programmes. The pattern of updating is often based on the individual requirements of the organization. In most cases, an end of the fiscal year summary file is produced which is maintained for a specified period of time (usually three years). In a few cases these data may be retained by the department for an extended period of time because collectively they can provide valuable trend data, the analysis of which could lead to more effective financial management practices within the organization. Because the amount of financial information managed by an organization is often substantial, financial data files are usually large in terms of the number of records involved. However, because most financial activities are centralized, each organization generally supports only one financial system.

3.30 In most cases, these data could be appraised as having informational value because of their high potential for secondary research use. In a limited number of cases a department might be legally required to maintain the financial data for a specified period of time. The data could therefore be appraised as having legal value.

3.31 The extent to which the data will be used will be determined by (in order of importance):

the nature of the financial activity;

the nature of the organization; and

the number of variables of information provided.

3.32 Their potential for secondary research use will also be dependent upon the degree to which they can be linked to other records. It should be expected that these data will be used for table production and some statistical (probably time series) analysis. They will rarely be used for record retrieval purposes. Although financial data, particularly those which simply document the financial status of an organization, have been traditionally treated as housekeeping records having little archival significance, it is useful to note that most departments employ financial researchers who depend upon the central financial data base for various research projects.

3.33 **Examples**

3.33.1 Financial data having a high potential for secondary research use normally describe regularly collected (annual) financial information covering, in considerable detail, the financial aspects of an activity which, in itself, is of substantial importance.

- financial statements for the government or organization as a whole
- savings bonds financial system
- loan assistance system for native persons

3.33.2 Financial data having less potential for secondary research use normally describe regularly collected (annual) or one-time financial information covering, in limited detail, the financial aspects of an activity which, in itself, is of limited national significance.

- Public Works contracts system
- National Research Council grants and scholarships financial system

3.33.3 Financial data having limited potential for secondary research use usually describe, in limited detail, the financial aspects of an activity which is of administrative significance only and would have little interest outside the administrative structure of the organization.

- departmental financial system

**Project Management Data**

3.34 These data describe information used by departments to monitor the progress of various activities or projects which are the responsibility of the department. These projects may exist entirely within the department or they may be conducted by outside interests on behalf of the department. The data may describe such variables of information as project name, project number, description of project, financial information, responsible officer, beginning date, target date, status reports, and so on.

3.35 The initial data describing the project and its supporting information are usually collected once with status reports and monitoring information added to the record as the project proceeds. Once the project is completed, all of the data pertaining to the project may be deleted. In a number of cases the data may be kept for a period of time for auditing purposes.

3.36 Dependent upon the amount of data gathered for each project, the records may be long and of variable length. Updating characteristics may be difficult to define because new data are often added on an irregular basis.

3.37 Because the number of projects supported by a department may be small, the overall size of many project management systems is often small. Most organizations, particularly those which centralize the administration of projects, support only one project management system.

3.38 These data may be appraised for either legal, evidential, or informational value. They may relate to a set of legal requirements affecting the department, or provide evidence of the overall administration of particular government programmes, or provide, to some extent, information on the areas affected by each project.

3.39 The extent to which the data will be used will be determined by (in order of importance):

the significance of the project activity; and

the nature of the organization.

3.40 Their potential for secondary research use will also be dependent upon the degree to which they can be linked to other records (for example, records describing similar projects or government programmes). It should be expected that these data will be used mainly for record retrieval purposes and to a lesser extent for statistical analysis and table production. In most cases, the greatest value of these data will be the evidence they provide of government activity in various projects.

3.41 **Examples**

3.41.1 Project management data having a high potential for secondary research use usually describe, in considerable detail, the activities of a large number of highly controversial projects administered by a government department.

- oil and gas well project management system
- Public Works building construction project management system

3.41.2 Project management data having less potential for secondary research use usually describe the activities of an ongoing research programme of a less controversial nature.

- Regional Economic Expansion projects system
- Supply and Services contracts data base

3.41.3 Project management data having limited potential for secondary research use frequently describe, in very limited detail, the status of non-controversial project activities that are generally administrative in nature.

- departmental internal research project accounting system

Operational Data

3.42 Operational data describe information collected and used by a department or agency to support activities conducted in accordance

with its legally established mandate. These data may have legal, evidential, or informational value.

3.43 The categories described in the following pages have been defined according to the types of data encountered by archivists in the Machine Readable Archives Division of the Public Archives of Canada. In this context they are only used to organize more clearly the narrative which follows. They are not designed to replace other definitions. They are as follows: measurement (or instrumentation) data; license data; survey data; registry data; and automated office information.

**Measurement (or Instrumentation) Data**

3.44 These are data collected by special instruments to measure a particular event. They can be taken at one time period or collected over time. They have short record lengths and are often voluminous and repetitive. They include measurements of specific rather than general events. They are often used to assist scientists or other researchers to understand better the event itself or to understand some process which is directly affected by the event. Examples of measurement data include oceanographic/bathymetric measurements, atmospheric measurements, geologic measurements, and so on. These data can cover a wide geographic area or they can be restricted to a small locality. Time spans can be very brief, very long, or one-time. Although there are exceptions, individual measurements contain only a few variables of information.

3.45 These data should be appraised as having informational value because of their high potential for research use. In nearly all cases they have some degree of potential research use. The extent to which the data will be used will be determined by (in order of importance):

the geographic area involved;

the time span; and

the number of variables of information provided.

3.46 Their research potential will also be dependent upon the degree to which they can be linked to other records.

3.47 **Examples**

3.47.1 Machine-readable measurement data having high potential for research use usually describe regularly collected atmospheric, bathymetric, geologic, and other similar measurements covering a wide geographic area (that is, national in scope).

- daily sea water measurements of coastal waters

- daily climatological measurements of the atmosphere
- a one-time geologic survey of a large area of a country

3.47.2 Machine-readable measurement data which might be considered to have less potential research use usually describe irregularly collected or one-time collections of atmospheric, bathymetric, geologic, and other similar measurements covering a region of a country (but they, too, must be national in scope).

- one-time series of geothermal measurements in permafrost regions
- oceanographic/bathymetric measures in the Labrador Sea, 1960-1964

3.48 Machine-readable measurement data which might be considered to have limited potential research use usually describe one-time collections of atmospheric, bathymetric, geologic, and other similar measurements covering a specific area of the country and/or related to a very specific research problem.

- one-time collection of geologic core data used to analyse stress loads for a proposed conference centre

**License Data**

3.48 These data describe information on licenses issued to an individual or organization performing an activity which must be licensed according to law. The data are frequently used to monitor or regulate these activities. They are usually collected on a regular basis (for example, when the license application or renewal is approved). Although individual records are often long, their number will vary according to the number of individuals or organizations involved. The data usually provide detailed information on the individual or institution, as well as the activities to which the license pertains. Collectively the license data are national in scope and not restricted to a particular area. In many cases they are maintained for the life of the license. Time spans, therefore, are dependent upon the life of the license. In many cases, however, the time span of the data should be at least one year. Because of their legal significance, some departments may be required to maintain the data for a specified period of time. In this case, the data may be appraised as having legal value. In most cases, however, the data should be appraised as having informational value because of their high potential for research use. In nearly all cases they will experience some degree of potential research use.

3.49 The extent to which the data will be used will be determined by (in order of importance):

the activity licensed;

the individuals or organizations licensed; and

the number of variables of information provided.

3.50 Their potential research use will also be dependent upon the degree to which they can be linked to other records. It should be expected that these data will be used for statistical analysis rather than record retrieval purposes.

3.51 **Examples**

3.51.1 License data having high potential research use usually describe regularly collected (annual) licensing information covering, in considerable detail, the activities of an important social group or group of nationally significant organizations.

- annual saltwater fisherman licensing system containing detailed information on individual fishermen, their vessels, species caught, ports visited, area fished, and violations
- annual radio licensing system containing detailed information on individual radio stations, their owners, staff, broadcast power and programming, geographic coverage, and so on

3.51.2 License data having less potential for research use frequently describe one-time license information covering the one-time activities of a group of individuals or organizations.

- one-time collection of license data describing the allowable extent of oil drilling activity in a specific area of the north
- one-time collection of license data containing information on the quotas and safety standards of a fleet of trucks carrying dangerous chemicals

3.51.3 License data having limited potential research use are usually irregular, regular, or one-time license data containing only a few variables of information.

- locomotive licensing system: annual updates of limited information describing the size, weight, type, and year of all locomotives operating in a country
- retail sale of fire-arms licensing system: updated every three years, containing limited information on the retail outlets involved in selling firearms (type of firearm, type of retail outlet location, owner, and sales information)

**Survey Data**

3.52 These data describe the results of research on a specific event or phenomenon affecting society. They may describe the opinions people have on events or phenomena, or they may describe the actual characteristics associated with particular events or phenomena.

3.53 The data are usually collected at one time but, in some cases, they may also be collected at regular intervals through time. Although they frequently describe the characteristics of specific rather than general events and phenomena, in the case of opinion surveys they may describe the attitudes of a general rather than specific populace to a specific event. Individual records contain a large number of variables. Individual data files, however, usually contain a relatively small number of records (for example, 1,000-3,000 records). Because surveys often represent a cross-section of either opinion information or the events and phenomena themselves, the complicated methodology as well as the sampling details must be properly documented.

3.54 Most survey data are placed in dormant storage or deleted once the department has completed its analysis. These data, therefore, are extremely dependent upon the factor of time; that is, the usefulness of the data for secondary research purposes may decrease with time, particularly if the event surveyed is extremely specific in nature.

3.55 Aside from the time factor, these data should be appraised as having informational value because of their high potential for secondary research use. In nearly all cases they will experience some degree of research use. The extent to which the data will be used will be determined by (in order of importance):

the significance of the event or phenomena being surveyed;

the degree to which the data will have significance for various research interests through time;

the degree to which they can be linked to other records or supplement other similar survey data files; and

the number of variables of information provided.

3.56 Survey data will normally be used for statistical analysis rather than record retrieval purposes. As many of these data already describe information which has been sampled, it is not recommended that a sample, extract, or summary of the data be considered for acquisition.

3.57 **Examples**

3.57.1 Machine-readable survey data having high potential for research use usually describe regularly collected (annual) survey information covering, in considerable detail, the characteristics associated with a particularly significant event or phenomenon.

- regularly collected survey data of a nationally significant demographic nature (census, crime, labour, drug abuse, and so on)
- regularly collected survey data of a non-demographic nature (such as industrial output, wages, construction, and so on)

3.57.2 Survey data having less potential for research use usually describe one-time survey and opinion survey information dealing with the characteristics associated with specific events or phenomena.

- one-time survey of a specific socio-economic event (such as a survey of opinions of a government programme)
- one-time survey of either a demographic or non-demographic nature which cannot be easily linked to other data files (such as a survey of physicians in 1972 or a survey of sawmill production in 1976)

3.57.3 Survey data having limited potential for research use usually describe, in limited detail, one-time survey and opinion survey information dealing with the characteristics associated with a very specific event or phenomenon.

- one-time survey of a very specific event which has little interest to potential researchers (such as a marketing survey of brand name products)
- one-time survey of a demographic nature covering a very unique group of individuals (such as a survey of government language course students in 1973)

**Registry Data**

3.58 These data are collected as the result of a legal requirement to register certain types of information. The data are normally used to administer and monitor a particular event. They are usually collected on a regular basis (that is, when the registration is recorded). Although the information on each record may be limited, the number of records may be large dependent upon the number of registered individuals or items. Collectively the registrations are national in scope and not restricted to a particular geographic area or local concern. In most cases they are maintained for the life of the registration. In some cases, particularly if an item must be registered on a regular basis, they are maintained beyond the life of the initial registration. The time span of each record, therefore, is dependent upon the length of time individuals or items must be registered.

3.59 Because of their legal significance, some departments may be required to maintain dormant registration data for a specified period of time. In this case, these data may be appraised as having legal value. In most cases, however, these data should be appraised as

having informational value because of their high potential for secondary research use.

3.60 The extent to which the data will be used will be determined by (in order of importance):

the activity registered;

the individuals or events being registered; and

the number of variables of information provided.

3.61 Their potential for secondary research use will also be dependent upon the degree to which they can be linked to other records. It should be expected that these data will be used for statistical analysis rather than record retrieval purposes.

3.62 **Examples**

3.62.1 Registry data having high potential research use normally describe regularly collected (annual) registry information covering, in considerable detail, the activities associated with an important social group or group of organizations or, collectively, a set of nationally significant events.

- central divorce registry containing demographic and other sociological background information
- chemical industry register containing financial and production details on all chemical industries
- aircraft accident register containing details on all aircraft accidents in the country
- passport registry
- taxation data

3.62.2 Registry data having less potential for research use normally describe, in limited detail, either regularly collected or one-time registration information covering the activities associated with a nationally significant group of individuals, organizations, or events.

- patents registry
- register of corporate names
- register of drug names and prices (1975)

3.62.3 Registry data having limited potential for research use normally describe, in very limited detail, one-time registration information covering the activities associated with a very specific group of individuals, organizations, or events.

- fishing licenses for a small geographical area

**Automated Office Information**

3.63 These data describe information often contained in correspondence, reports, memoranda, and other documents which are stored in a machine-readable form (such as word processing diskettes). Although they are frequently used to speed the editing and production of textual or paper documents, they may also be used exclusively in a machine-readable form through a telecommunications network linking a number of terminals.

3.64 The data are usually created at one time, edited over a brief period of time, and then deleted after the paper product is produced. In a number of cases, however, they may be stored on diskettes over a longer period and, although edited versions may be produced occasionally in paper form, the initial format may reside on disk to satisfy a number of related requirements. The records are usually full text and may vary in length depending upon the type of information recorded. Some records may be as brief as a short memorandum, while others such as mailing lists and large reports may be longer.

3.65 The usefulness of these records over the long term may be limited as in many cases retention of the paper copy of the machine-readable information would suffice. In a number of cases, however, these records should be considered for archival retention because their usefulness may be greater if they are stored in machine-readable rather than paper form.

3.66 **Examples**

3.66.1 Machine-readable automated office data having high potential for future use usually describe information which is particularly significant to the operations of the institution both from an administrative and historical perspective.

- ministerial records

3.65 The usefulness of these records over the long term may be limited as in many cases retention of the paper copy of theless potential for future use usually describe information contained in memoranda, correspondence, and other documents considered to have less significance to the organization than those described above.

- intra-office memoranda
- correspondence

3.66.3 Machine-readable automated office data having limited potential for future use usually describe information contained in published reports, lists, and other documents which are widely available and which have limited potential for either manipulation or linkage to other documents.

- annual reports

## SUMMARY

3.67 The appraisal of machine-readable records involves the evaluation of the information in the records (content analysis), as well as an evaluation of the technical aspects of the records (technical analysis). The first part of this chapter has described the procedures involved in undertaking the content analysis of machine-readable records. It has been pointed out that these are not unlike the traditional activities of archival appraisal which require an examination of the evidential, informational, and legal value of records. Each of these factors has been discussed briefly, emphasizing those features which are particular to machine-readable records.

3.68 In order to assist in the application of informational, evidential, and legal values to machine-readable records, a more detailed analysis of these attributes has been provided in the second part of the chapter as they relate to individual categories of information. The categories have been developed in a somewhat arbitrary fashion, according to the subject classes of information most commonly managed in machine-readable form. Administrative (or housekeeping) data have been divided into four categories: personnel; supply; financial; and project management. Operational data have been divided into five broad categories: measurement (or instrumentation); license; survey; registry; and automated office information. Each of these categories have been further sub-divided according to the relative research values of various types of information associated with each.

## NOTES

1. Frank B. Evans, et. al, "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers," The American Archivist, Volume 37, Number 3, July 1974, page 417.

2. Theodore R. Schellenberg, The Appraisal of Modern Records. Washington: National Archives, 1956, page 45.

3. Ibid., page 11.

4. Meyer H. Fishbein, "The Evidential Value of Nontextual Records: An Early Precedent," The American Archivist, Volume 45, Number 2, Spring 1982, pages 189-190.

5. Thomas Brown, The Impact of the Federal Use of Modern Technology on Appraisal, an internal report prepared for The Appraisal and Disposition Task Force of the National Archives and Records Service, Washington, D.C., March 1983, pages 74-75.

6. For a more detailed explanation of the procedures involved in aggregation, see Chapter V, paragraphs 5.15 to 5.18 inclusive.

## Chapter IV

## APPRAISAL OF MACHINE-READABLE RECORDS: PART II

4.1 The appraisal of the contents of machine-readable records follows, with some exceptions, the same guidelines as the appraisal of traditional records. However, machine-readable records cannot be appraised solely for their content. They must also be examined in terms of their technical requirements. An obvious example is that if the machine-readable records are damaged and are unable to be read, they cannot be appraised as having research value. It is also possible that the internal arrangement of the data could be so complex that it might affect the processing of the records and the eventual servicing of the records. The size of the machine-readable data files and how often they are updated could certainly be an important consideration in the future preservation costs of the records. In theory perhaps such technical considerations should not play a role in the appraisal process but, due to the vulnerability of the present storage medium and the costs incurred in the processing and servicing of the records, these factors cannot be ignored.

4.2 The archivist must be aware of how the machine-readable records being appraised came into existence, of other similar records which exist in other formats, and the methods and approximate costs which will be used in properly conserving the information. These questions require an in-depth knowledge of the records as well as any related records. Some knowledge of data processing techniques and methods used in the field is also required in order to evaluate the technical aspects of the records.

### TECHNICAL ANALYSIS - AN OVERVIEW

4.3 Before any judgment can be made on the historical or long-term research value of machine-readable records, it must first be determined if the computer tape can be read. While computer tape remains the most common storage medium for machine-readable records, other kinds of physical media exist (including punched cards, punched tape, magnetic disks, floppy disks, and magnetic drums).

### Readability of the Records

4.4 This involves checking the physical readability of the magnetic tape by placing the tape on a tape drive and instructing the computer to read the information. If this is not possible, then the archivist must try to determine the last time the tape was read, any problems which were encountered, where the tape has been stored, and so on.

4.5 Sometimes temporary read errors are present which can be eliminated by passing the tape through a tape cleaner. Should permanent read errors be encountered, then the decision about readability will depend upon both the scope and magnitude of the errors. If only a few blocks are unreadable, this may not seriously affect the value of the file. However, if more than five per cent of the records cannot be read, then in most cases the machine-readable data file would be considered unreadable. At the same time that a tape is read, it is a good policy to obtain a printout or "dump" of a few records from the data file. This will enable the archivist to determine if the documentation corresponds with the data.

4.6 Data obtained on other storage media should also be checked for readability. This should be undertaken by testing it on an appropriate input device.

Adequacy of the Documentation

4.7 This refers to the textual documentation which accompanies machine-readable records. The minimum requirement is a record layout and a codebook. The record layout is a diagram or a list of the contents of a logical record which describes the item of information in each field, the length of each field, and the position of each field in the record. The codebook provides an explanation of the codes used to represent information in numeric or abbreviated form. Minimum documentation must be looked at in terms of sufficient information to appraise and process the records and sufficient information for a researcher to use the records in secondary research. In some cases, the amount of documentation required for all three uses will be the same. In other cases, a record layout would not be sufficient for secondary analysis. An archivist should review the documentation with all uses in mind.

4.8 It is possible that the necessary documentation can be located in the same file, a user's guide, or some other collection of material. However, in many cases the archivist will find the documentation in separate locations within a department, or even in separate departments. It is imperative that sufficient documentation to appraise, process, and use the records be located, as otherwise the machine-readable data cannot be used.

4.9 If the data can be read and there is adequate documentation, then the archivist can proceed to the analysis of the contents of the machine-readable records (as described in the preceding chapter), and a more detailed technical analysis of the arrangement of the records and problems which could occur due to long-term storage (as outlined below). The following steps involve a detailed examination of the technical requirements of the records in terms of the implications of such requirements on the acquisition, processing, preservation, servicing, and restrictions on access.

## Acquisition and Processing Implications

4.10 The archivist should examine the size of the file, its complexity in terms of its internal structure, and its degree of dependence on specific software and hardware. If the size of the file is in question, the archivist might have to consider the possibility of obtaining only a sample of the records. Should this be the case, then the effect sampling might have on the informational value would have to be determined.(1)

4.11 The internal arrangement of the data must be examined in order to determine the implications that this will have on the processing of the file. Processing refers to the act of comparing or verifying sample printouts of the data with the textual documentation containing the record layout and codebook.

4.12 The actual arrangement of the individual records on the reel of tape is seldom a major consideration as it is a fairly simple and inexpensive process to place the records in any specified order. However, one file might use an unusual character code, while another might have a standard character code. One file might be used only with certain computer programs, while the other might operate with any program. Or, one file might require some unusual piece of hardware equipment made by a specific manufacturer, while the other might be processed on almost any equipment. In most cases the file that uses the standard codes, a number of different programs, and common equipment will be considered to have the more desirable arrangement. For those machine-readable records containing information duplicated by textual records, the machine-readable records will, in the majority of cases, be appraised as having the better arrangement because they have greater manipulability.

4.13 A major consideration when undertaking the technical analysis is the hardware dependency of various storage media and the software dependency of certain formats of information. Hardware dependency is when processing the information is dependent upon a specific piece or type of equipment, while software dependency is when processing the information is dependent upon a specific set of computer programs. It must be pointed out, however, that these types of dependencies are not that distinct. Some types of physical equipment are dependent upon certain software. Conversely, some software can be used only with certain kinds of equipment.

4.14 When appraising machine-readable data files which are software dependent, it is important for archivists to "consider the costs associated with reformatting the data or acquiring the necessary software as well as the potential for reducing the utility of the data by reformatting it." There will be occasions when the software requirements are such that the archival repository may be unable to acquire a data file. Needless to say, reformatting data

into a software dependent form or acquiring software needed to access the data will greatly increase the cost of accessioning a data file.(2)

4.15 There will also be occasions when it is necessary to reformat machine-readable data files because they are dependent upon particular computer equipment or hardware. Indeed, it may be found that some data files require rare or obsolete hardware in order to be processed. In such situations, it may be necessary to have the data processed on the same machine which originally created the digitized information and which may still be in the possession of the creating institution. The second step will be to produce the information in an intelligible format onto a medium without hardware dependency problems.

4.16 In those archival repositories which have dealt with machine-readable records to date, it has been the practice to accession only hardware and software independent files. However, this has posed problems, and continues to do so. It has presented a number of choices to the repositories. First of all, they can attempt to make it a requirement for the creating institutions to provide the repository with only hardware and software independent data files. Secondly, they can decline to accession the information in its dependent format(s) and authorize its destruction, with the result that important information will be lost. Thirdly, the repositories can reformat the information themselves, realizing that this option obligates them to expend resources to reformat the information into a usable form. And finally, the repositories can accession the data as they exist and leave it to researchers to cope with the technical problems. This alternative may leave the information in such an unusual structure that the archivists are unable to determine what information has been transferred and therefore unable to describe the contents of the records and to assist researchers in accessing the data.

## Preservation Implications

4.17 The implications on the preservation of machine-readable records involves the cost of preserving the data. Preservation costs are usually determined according to the number of reels of magnetic tape. Included in the analysis would be the following: the purchase price of a reel of magnetic tape; the cost of recopying the tape as well as the frequency of the recopying programme; and whether a backup copy of every tape is maintained. If the informational content of a file is marginal and the file consists of several hundred reels of tape, it is possible that the file would be rejected because of excessive preservation costs.

**Servicing Implications**

4.18 The complexity of a machine-readable data file and the format in which the information will be preserved might have implications on the servicing of the file. It is also essential to determine who will service the file-whether the originating department or the archival repository. If the file is software and hardware dependent, it might be easier for the originating department to handle all service requests. As the archivist has already appraised the content of the records, including their current and potential research use, it should be possible to determine, in consultation with the creator, how the data would be used.

**Restrictions on Use**

4.19 The same kind of restrictions apply to machine-readable records as to textual records. However, the manner in which such restrictions are handled is different. If only certain portions of information are restricted, it is possible to remove all other portions of information from the file for research use, thereby creating a public use version of a restricted file. However, many machine-readable data files contain information on individuals and have a restriction which prohibits the release of an individual's identity. Removing only the personal identifiers may not protect the individual's identity, as the specific person might be identified from the remaining information in the file. Therefore, in weighing the restrictions on a file, the archivist must consider what specific portions of information could or should be withheld to prevent disclosure of restricted information. After this has been determined, the archivist must then consider the impact of these restrictions on the informational value, as well as the cost of producing a public use version.(3)

**TECHNICAL ANALYSIS - APPLIED TO SPECIFIC KINDS OR TYPES OF DATA**

4.20 This section has been designed to assist in the evaluation of such technical attributes as software, hardware, size, and physical arrangement for various types or kinds of data. Each data type has been defined according to a set of common technical characteristics. The major characteristic affecting these definitions is the software environment within which each data type was originally managed (such as data base, numeric or statistical, cartographic, textual, and so on). Only three kinds or types of machine-readable data are assessed - survey files, time series files, and data bases. The approach can certainly be applied to other data types. However, it is with these types of data that archivists have had more experience to date. As archivists gain more knowledge and experience with digitized cartographic and textual data, the approach could and should be expanded and/or modified.

4.21 An analysis of the technical considerations of the file should lead to a more rational development of an approach to the acquisition,

processing, and servicing of the data. The approach itself should be developed according to the willingness of an archival repository to absorb the costs associated with each of these archival functions. To this end, and to permit a more systematic analysis of these considerations, a set of question - and - answer planning tools are included with the definitions of each data type. These forms, however, are only a guide and should not be used without careful consideration to all of the unique characteristics which will undoubtedly be associated with each machine-readable data file.

4.22 For each data type, the last question in the question-and-answer form sets the stage for the Evaluation of the results of the Technical Analysis associated with a file as measured against the results of the Content Analysis. The results of the Evaluation should lead to the decision to either accept or reject the file.

4.23 In most cases, a file appraised as acceptable should be acquired immediately. In some cases, however, the decision to acquire may be postponed. The operational life of the data, for instance, and the willingness of the creating department to assist in the servicing of a data file, could affect the timing of its acquisition and even its archival format.

4.24 Outlined below are only a few of the possible strategies that could be developed for a number of such situations. It is important to note that the appraisal decision and the servicing conditions in each example could affect the final archival format of the files (for example, a system file for X number of years, and a sequential file thereafter). All are based on the existence of retention periods which must be applied to all files prior to the appraisal process. Although a number of retention periods may exist to account for the various processes (transactions, initial data, and so on) of a major computer system, the following examples are related to master files only.

4.25 **Type One Data File:** One-time data colection; no data are added.

**Example 1:**

| | |
|---|---|
| Suggested Retention Period: | immediate disposal (department to scratch the data file immediately after creation and analysis). |
| Appraisal Decision: | acceptable file; a copy of the data file to be transferred to the archives immediately after creation. |
| Servicing Conditions: | the archives will handle all requests. |

**Example 2:**

Suggested Retention Period: ten years dormant and scratch (department will copy the data to dormant storage tapes immediately after creation and analysis).

Appraisal Decision: acceptable file; a copy of the data file to be transferred to the archives immediately after creation.

Servicing Conditions: requests on the data file will be redirected to the creating department for a period of ten years, after which all requests will be handled by the archives; agreement to be renewed immediately prior to the deletion of the dormant storage tapes.

4.26 **Type Two Data File:** Continually updated; all historical data and all updates are actively maintained in the file (no data are deleted or copied onto storage tapes).

**Example 1:**

Suggested Retention Periods: ten years and review [department will actively maintain all data in the data file for ten years; that is, no data will be scratched or copied onto storage tapes; records management, archives, and departmental personnel will review the status of the data file in ten years, at which time the schedule will either be renewed (that is, another ten years and review), or a more specific set of retention periods could be applied].

Appraisal Decision: acceptable file but the archives will not acquire any portion of the data file until the department indicates a disposal date through a revised schedule.

Servicing Conditions: any requests for the data will be redirected to the department.

4.27 **Type Three Data File:** Continually updated; historical data are deleted or copied onto storage tapes in order to keep the file current.

**Example 1:**

Suggested Retention Period: earliest full year of data to be scratched in ten years (department will continually and actively maintain the ten most recent years of data as part of the data file; no data will be scratched or copied onto storage tapes; earliest year of data will be scratched as the most recent year of data is created).

Appraisal Decision: acceptable file; on an annual basis, the archives will acquire a copy of each year of the data file as it is created.

Servicing Conditions: requests on most recent ten years of data to be redirected to the department; requests on data created prior to the most recent ten years of data to be handled by the archives; agreement to be reviewed immediately prior to the department's first deletion of its earliest year of data.

**Example 2:**

Suggested Retention Period: ten years, earliest full year of data dormant five years (department will continually and actively maintain the ten most recent years of data as part of the data file; that is, no data will be scratched or copied onto storage tapes; earliest full year of data will be copied onto storage tapes and maintained in dormant storage for five years and scratched).

Appraisal Decision: acceptable file; on an annual basis the archives will acquire a copy of each year of the data file as it is created.

Servicing Conditions: requests on most recent fifteen years of data to be redirected to the department; requests on data created prior to the most recent fifteen years of data will be handled by the archives; agreement to be reviewed immediately prior to the department's first deletion of its earliest year of data.

**Example 3:**

Suggested Retention Period: active for operational life of individual records (for example, length of time individual receives benefits, and so on), dormant five years (department will copy data pertaining to individuals onto storage tapes after the operational life of their data is over; to be held in dormant storage for five years and scratched).

Appraisal Decision: acceptable file; the archives will acquire a copy of the dormant storage tape on an annual basis (if the department had preferred to scratch individual records after their operational usefulness, the archives could have requested that these data be copied to tape for acquisition on an annual basis or, alternatively, requested a snapshot of the data file on a regular basis).

Servicing Conditions: requests on data created prior to the most recent five years of data will be handled by the archives (if the department had preferred to scratch individual records after their operational usefulness, the archives could have handled all requests for historical data; prior to responding to the request, however, the archives staff would have had to ensure that the record was not active; if it was active, the request would have been redirected to the department).

4.28 Survey files are machine-readable data files containing collections of information relating to a person, thing, or activity. Surveys are usually undertaken to describe the characteristics of a population, to evaluate programmes, to generate explanatory models, or to test hypotheses. Surveys are most common in the field of social science. In machine-readable form they are often managed by statistical packages such as SPSS, SAS, OSIRIS, DATATEXT, and so on.

4.29 Surveys can be of three types:

one-time survey: a population is surveyed at one specific point in time.

example: National Election Study, 1974

longitudinal: a population is surveyed at different points in time but, although the population may have the same characteristics, the people are not the same.

example: Trends in Drug Use Among Rural Students, 1974

panel survey: the same population is surveyed at several points in time and the people included in the population are always the same. Panel studies are often used to study trends.

example: Panel Study of Income Dynamics

4.30 Essential elements of survey files are as follows:

questionnaire design;

carefully specified hypothesis;

proper sampling procedures; and

careful data capture and entry.

4.31 The inclusion of certain basic variables will certainly increase the usefulness of the data and likely increase their value and their compatability with data created from other sources of information. The question of anonymization is most important with respect to survey files; that is, it will have to be determined if the data need to be anonymized and, if so, what effect it will have on the usefulness of the data.

4.32 The technical requirements must be assessed in two ways: sampling methodology; and the format of the data.

4.32.1 Sampling Methodology

The archivist should have appraised the sampling methodology at the same time as the contents of the records. It is an integral part of the determination of informational value.

4.32.2 **Format of the Data**

The format in which the data have been coded is the main question in appraising the technical requirements of the file. To provide the researcher with the basic information from which he/she can work, it is more useful to be able to acquire the basic raw data. Such data can be found in either a system file using one of the statistical packages, or simply the raw data. If the archivist is appraising a system file, it is necessary to ensure that the raw data file is still intact and accessible. Creators of system files often create their own variables and perform a number of recodes to their own specifications. Such changes could substantially affect the use which could be made of the file by someone else.

4.33 The technical considerations of survey files are not as far - reaching as with data bases. However, they must be evaluated carefully. The interrelationships between variables and/or records will certainly play a significant role in the verification of the data and therefore in the cost of processing the data.

4.34 The Planning Tool which is described below could be used to undertake a full evaluation of the file and to provide some indication of the implications for the archival repository of acquiring, processing, preserving, and servicing a particular file.

4.35 **Survey File Planning Tool**

4.35.1 Is this a survey file?

_____Yes

_____No

4.35.2 If Yes, which type of survey file is it?

_____One-time survey

_____Longitudinal survey

_____Panel survey

_____Other (specify):_____

4.35.3 In what format are the data?

_____Raw data

_____System file

4.35.4 (If system file) Which statistical package was used?

_____SPSS _____DATATEXT _____

_____OSIRIS _____BMD _____

_____SAS _____P-STAT _____

_____Other:(specify):_____

4.35.5 What kind of hardware supported the package?

_____Burroughs _____Honneywell _____Xerox

_____CDC _____IBM

_____PDP _____UNIVAC

_____Other(specify): _____

4.35.6 If the data are in a system file, have there been a substantial number of records and variables created?

_____Yes

_____No

4.35.7 If Yes in 6, can the raw data be accessed using the system file?

_____Yes

_____No

4.35.8 If no in 7, can the data be converted to raw format?

_____Yes

_____No

4.35.9 Number of variables: _____

4.35.10 Number of logical records: _____

4.35.11 Are there personal identifiers contained in the file?

_____Yes

_____No

4.35.12 Is it possible to create an anonymized version of the file without changing the data substantially?

_____Yes

_____No

4.35.13 How much will it cost the archives to manage the data?

a) Acquisition Costs (tape copy/documentation) — computer costs $_____ ; person/time (dys) _____

b) Processing Costs (conversion, if required/ verification/archival copies) — computer costs $_____ ; person/time (dys) _____

c) Conservation Costs (per year) (tape storage/rewinding/ recopying) — computer costs $_____ ; person/time (dys) _____

d) Servicing Costs

i) anonymization — computer costs $_____ ; person/time (dys) _____

ii) service to public (tape copy/documencation) — computer costs $_____ ; person/time (dys) _____

Evaluation

4.35.14 Is the archives prepared to pay the costs to manage the data according to the proposed format?

_____Yes (complete appraisal form)

_____No (reconsider value of data and/or archival format)

Total Computer Costs $_____
Total Person Costs $_____
Total Person/Time (dys) _____

## Time Series Files

4.36 Time series can be defined as the discrete or continuous sequence of quantitative data assigned to specific moments in time, usually studied with respect to their distribution in time. In time series files the elements of observation are chronologically recorded and the true "records" in the file are represented by variables or series. Time series data are usually updated regularly (daily, weekly, monthly, and so on) and are often associated with the field of economics. They can also be used in relation to measurement re-

cording. Time series files are simple in the format used as compared to the same type of data managed by a data base management system. The data are often managed by such packages as MASSAGER, MATOP, DATABANK, and so on.

4.37 There are a number of important elements which must be known in order to appraise the technical aspects of time series files: the number of variables or series; the average number of logical records; the frequency of the updating of the records; and whether or not data are scratched during the updating of the files. The archivist should also investigate if there will be a termination date for the series.

4.38 In order for data to be classified as time series, there must be no change in the variables defined for which data are collected.

4.39 Listed below are a number of examples of time series files:

Industry Selling Prices (containing data describing industry selling prices of a variety of commodities including foods, manufactured goods, clothing, building materials, and vehicles, the data being seasonally adjusted and unadjusted);

Weekly Food Prices Monitoring Program (containing data on weekly price indices calculated from the data of eight cities).

4.40 **Time Series Files Planning Tool**

4.40.1 Is this a time series file?

_____Yes

_____No

4.40.2 Which package was used to manage the data?

_____ MASSAGER _____TROLL

_____MATOP _____Other(specify):_____

_____DATABANK

4.40.3 What hardware supported the package?

_____Burroughs _____Honeywell Xerox_____

_____CDC _____IBM Other(specify):_____

_____PDP _____UNIVAC

4.40.4 How many series are contained in the file?_____

4.40.5 What is the estimated size of the file?
(average # of logical records) _____

4.40.6 How frequently are the series updated?

_____Daily _____Monthly Yearly_____

_____Weekly _____Quarterly Other(specify):_____

4.40.7 Does the department/agency know how long the series will be updated?

_____Yes (provide date): _____

_____No

4.40.8 Are data or series deleted from the files?

_____Yes

_____No

4.40.9 At which point in time could the updates be acquired?_____

4.40.10 How much will it cost the archives to manage the data?

a) i) Acquisition Costs (tape copy/documentation) — Computer costs $_____ person/time (dys) _____

ii) Projected Acquisition Costs for Updates — computer costs $_____ person/time (dys) _____

b) Processing Costs (conversion, if required/ verification/archival copies) — computer costs $_____ person/time (dys) _____

c) Conservation Costs (per year) (tape storage/rewinding/ recopying) — computer costs $_____ person/time (dys) _____

d) Servicing Costs — computer costs $_____ person/time (dys) _____

**Evaluation**

4.40.11 Is the archives prepared to pay the costs to manage the data according to the proposed format?

_____Yes (complete appraisal form)

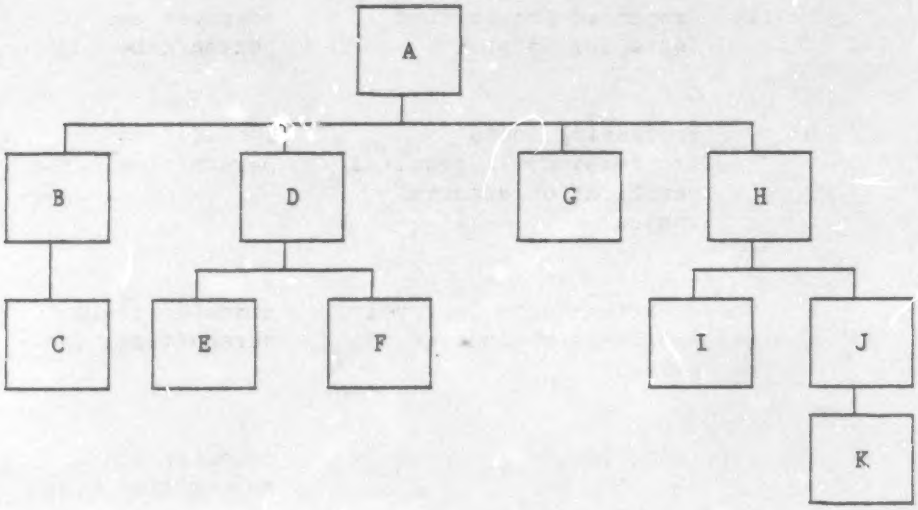
_____No (reconsider value of data and/or archival format)

Total Computer Costs $_____
Total Person Costs $_____
Total Person/Time (dys) _____

**Data Bases**

4.41 A data bases is a complex arrangement of data which is dependent upon a sophisticated set of software known as a data base management system (DBMS).

4.42 Data bases go one step further than files in terms of data organization. Whereas data files collect bits into characters, characters into data items, and items into records, data bases do the above, plus collect records into sets. A set is defined as a collection of records that are related in some way. Each set type has a unique name and any number of sets of a given type can exist. However, each set type is defined independently of all other set types. A set type is a two-level structure which includes owner record types and member record types. In a data base environment, these are often assembled into either a hierarchical or a network structure. The chaining of each of the record types into a set is accomplished through the use of pointers, often imbedded within the data themselves. Thus, in the figure provided below which is an example of an hierarchical structure, a series of pointers might direct the access of "A" to either one of "B", "D", "G", or "H", depending upon the set required.



4.43 The record types of the next level would, in turn, contain pointers to various other branches, and so on down the hierarchy. In this way, one record type could become the owner of several record type members. This concept of access through pointers is considered an important data base feature. It is also the reason why all data bases are managed in a disk environment.

4.44 Another important feature which evolves from the foregoing discussion is the concept of data independence. That is, the entire data base does not have to be accessed in order to respond to an application requiring only a small segment of the data. A data base, therefore, is usually designed so that only one part needs to be accessed by a particular program for a given application.

4.45 The software used to define and manipulate a data base is known as a data base management system (DBMS). A DBMS is necessarily complex in structure and, as a result of enhancements or technical developments, can be subject to numerous modifications in its specifications over time.

4.46 A data base, therefore, is dependent upon three things: the complex physical structure associated with its design; the DBMS supporting the data base; and sometimes the hardware environment in which it is managed.

4.47 The data base management systems most commonly used are:

| | | | | |
|---|---|---|---|---|
| ADABAS | DMS 1100 | IDMS | MANAGE | RAPID |
| BASIS | DMS 11 | IMAGE/3000 | MARSVI | SIR |
| COMPOSIT-77 | EDMS(XEROX) | IMS/VS | MIDAS | SYSTEM 1022 |
| DBMS 11 | IDS/11 | INQUIRE | MPLDR | SYSTEM 2000 |
| | | | | TOTAL |

4.48 Data base management systems manage information in three different ways. One type of system retains all data even after updates are made to individual records. Therefore, these DBMS do not have a purge routine for non-current information. The second type are those systems which substitute updated information for non-current information. These purge routines can either delete the information entirely or store the information in an historical file, generally on tape. The last type are the data base management systems which deal with one-time data collections and no information is ever updated.

4.49 There are usually three different formats in which an archives may accept information from a data base. These are a system dump, an unloaded character string, or a sequential file structure. Each option poses certain problems and these must be known if the archivist is to undertake a complete appraisal.

4.50 A system dump is a dump of the DBMS file exactly as it appears on the disk of the DBMS. Such an output has all of the problems associated with any disk dump, including dependency of the operating system which produced the disk. It is also dependent upon the particular version of the DBMS and the particular type of hardware which created the format. The major advantage of this process is that it is relatively inexpensive to dump a disk and reload it onto the same machine.

4.51 The second option, of acquiring an unloaded character string output onto tape, can be provided by those data base management systems that permit the data to be "unloaded" from DBMS onto tape in a usable format. In this arrangement, each individual file in the hierarchy is unloaded separately. While having a sequential arrangement of records in every file, the records of such an unloaded data base are variable in length, the field identifiers are imbedded, and several data fields or record types may be repeated, depending upon the requirements of the individual records. Individual fields may also be variable in length. But, because the field identifiers are included and a logical pattern exists to the information, data in this form can be processed. It must be kept in mind, however, that these routines were not developed primarily or necessarily to process the information with other software. Therefore, while it is theoretically possible to process the unloaded character string output with other computer programs, for all practical purposes the data can only be easily accessed by using a version of the same DBMS. Because of this, the archival repository or a researcher would require access to a version of the DBMS which created the original information in order to use the information.

4.52 The third option, reformatting the data into sequential fixed - length records, uses a sophisticated program to copy each set of data onto a tape with progressively more data being included on the tape as data sets further down the hierarchy are copied. To refer to the diagram in paragraph 4.42 above, this means that file A would be copied, then file A-B, then file A-D, then file A-G, then file A-H, then A-B-C, and so on down the hierarchy. Another option would be to group several data files together as groups of files. The most important advantage to this option is that the information would be independent of the original hardware and DBMS. Therefore, any programming language on any machine could process the data. There are, however, a number of serious disadvantages. First of all, the programming involved in converting the data base structure to a sequential format is both time-consuming and costly. Both the person/time and computer costs would have to be repeated whenever a data base was acquired. Secondly, the number of data repetitions and data files involved would make it difficult for a researcher to access and retrieve selected portions of information as the information would be in an awkward form for retrieval.(4)

4.53 **Data Bases Planning Tool**

4.53.1 **Is this a data base?**

_____Yes

_____No

If no, are the data managed by one of the following data base management systems (DBMS)?

| | | | |
|---|---|---|---|
| ___ ADABAS | ___ EDMS | ___ INQUIRE | ___ RAPID |
| ___ BASIS | ___ IDS II | ___ MANAGE | ___ SIR |
| ___COMPOSIT-77 | ___ IDMS | ___ MARS VI | ___ SYSTEM 1022 |
| ___ DMS 1100 | ___ IMAGE | ___ MIDAS | ___ SYSTEM 2000 |
| ___ DMS II | ___ IMS/VS | ___ MPLDR | ___ TOTAL |

NOTE: If none of the above, reconsider the data type by referring to other categories.

4.53.2 What kind of analysis was originally performed on the data? (indicate % of time for each)

___ ___% occasional record retrieval
___ ___% occasional record linkage
___ ___% occasional table production
___ ___% occasional statistical analysis

4.53.3 What kind of analysis will the research community be expected to perform on the data? (indicate % of time for each)

___ ___% occasional record retrieval
___ ___% occasional record linkage
___ ___% occasional table production
___ ___% occasional statistical analysis

4.53.4 Will the entire data base be required for the purposes outlined in 3?

_____Yes (go to 6)

_____No

4.53.5 If no, what part of the data base would suffice for the analysis or retrieval purposes indicated in 3?

___ whole

___ extract

___ summary

___ aggregate

___ sample

4.53.6 How many times will the research community be expected to access this data base/part of the data base?

___ less than once per year

___ 1-5 time per year

___ 6-12 times per year

___ more than 12 times per year

4.53.7 What kind of software originally managed the data base?

| | | | |
|---|---|---|---|
| ___ ADABAS | ___ EDMS | ___ INQUIRE | ___ RAPID |
| ___ BASIS | ___ IDS | ___ MANAGE | ___ SIR |
| ___ COMPOSIT-77 | ___ IDMS | ___ MARS VI | ___ SYSTEM 1022 |
| ___ DMS 1100 | ___ IMAGE 3000 | ___ MIDAS | ___ SYSTEM 2000 |
| ___ DMS II | ___ IDMS/VS | ___ MPLDR | ___ TOTAL |

4.53.8 What kind of hardware originally supported the data base?

| | | |
|---|---|---|
| ___ Burroughs | ___ Honeywell | ___ Xerox |
| ___ CDC | ___ IBM | ___ Other(specify): _____ |
| ___ PDP | ___ UNIVAC | |

4.53.9 Would most researchers have reasonable access to a computer center supporting the hardware and software described above?

___ Yes

___ No (convert data to a sequential raw format)

4.53.10 Would the archives have access to a computer center supporting the hardware and software described above?

___ Yes

___ No (convert data to a sequential raw format)

4.53.11 Given the information described above, what are the most appropriate archival formats for the data?

___ system dump

___ sequential unload (packed, variable length)

___ fixed length sequential

4.53.12 For what length of time will the current research community be expected to use the data in the archival format described above?

___ less than one year (review date_____)
___ 1-5 years (review date _____)
___ 5-10 years (review date _____)
___ 10-15 years (review date _____)
___ 15-20 years (review date _____)
___ 20-25 years (review date _____)
___ 25-30 years (review date _____)
___ more than 30 years (review date _____)

4.53.13 How much will it cost the researcher to restore and use the data base according to the archival format described above? (calculate per annum)

computer costs:

___ less than $50
___ $50-$100
___ $100-$500
___ $500-$1,000
___ more than $1,000

4.53.14 Will the research community be prepared to pay the amount described above?

___ Yes

___ No (reconsider archival format)

4.53.15 How much will it cost the archives to manage the data according to the format described above?

a) Acquisition Costs (tape copy/program preparation) — computer costs $_____ ; person/time (dys) _____

b) Processing Costs (documentation preparation/ data checks) — computer costs $_____ ; person/time (dys) _____

c) Conservation Costs (per year) (tape storage/rewinding/ recopying) — computer costs $_____ ; person/time (dys) _____

d) Servicing Costs (per year) (tape copy/photocopies/data base restoration/ programming) — computer costs $_____ ; person/time (dys) _____

Evaluation

4.53.16 With respect to the archival value of the data, is the archives prepared to pay the costs of managing the data base according to the proposed archival format, and for the period of time up to the first archival review date?

___ Yes (complete appraisal form)

___ No (reconsider value of data and/or archival format)

Total Computer Costs $_____

Total Person Costs $_____

Total Person/Time (dys) _____

## EVALUATION AND PLAN OF ACTION

4.54 It is at this stage that the archivist should bring together the results of the Content Analysis and Technical Analysis and justify the decision to acquire the records or to reject the records. In many cases the evaluation should be fairly straightforward. For example, a file may have such high research value and require such

little cost in terms of acquisition, processing, etc. that a recommendation to "accept" can be easily justified. Similarly, a file that has extremely limited research value and could potentially absorb substantial costs in terms of acquisition and processing could easily be recommended for rejection.

4.55 In other cases, however, the justification to accept or reject could be more difficult to develop. Weighing the research value of a file against the costs of its acquisition, processing, preservation, and servicing may require the assistance and expertise of not only the archivist and his/her supervisor, but also the users in the creating institution as well as any research groups which might have a potential interest in the file.

4.56 In most cases, a file appraised as "acceptable" should be acquired immediately. However, in some cases the decision to acquire may be postponed. The operational life of the data, for example, and the willingness of the creating institution to assist in the servicing of the data, could very well affect the timing of acquisition and even the archival format.

4.57 Should the appraisal decision be favourable, it is then suggested that a plan of action be developed, using the information contained in the planning tools. The action plan should cover the acquisition, processing, conservation, and servicing functions.

Acquisition Plan

4.58 The acquisition plan should outline the steps that will be followed to acquire the file from the donor agency. The format of the data, size, number of tapes required, programs required, computer centers involved, and the general roles of the archival repository and the donor agency should be described. The amount and types of documentation and their method of acquisition (photocopying, microfilm, direct transfer, etc.) should also be indicated. If the file is not to be acquired immediately, then the possible acquisition date should be indicated and a possible acquisition strategy should be provided.

4.59 If at all possible, the approximate costs of the acquisition should be calculated (such as tape copying, program preparation, and so on). The amount of staff resources (both time and salary dollars) would also be useful. Initially, archivists will not be able to provide such information but, as more experience is gained in the archival handling of machine-readable records and if performance measurement data are collected and compiled, then the inclusion of such information should be possible. This same comments holds true for all of the archival functions described below.

**Processing Plan**

4.60 There should be a description of the suggested processing strategy and an explanation of the steps that should be taken in processing the acquired file into an archival format. Verification checks (label, dump, frequencies) as well as the degree of verification (partial dump only, all data items verified, etc.) should also be described. Documentation preparation and arrangement, including the degree to which the documentation should be converted to an archival standard, should also be included.

4.61 The final archival format of the data should be described in general terms. This should include size (number of characters, number of records, etc.), arrangement (sequential, system dump, etc.), complexity, number of tapes required, software dependency (if any), and hardware dependency (if any).

4.62 The approximate cost of processing the file should be calculated, if possible. This would include such things as computer processing, conversion, verification, documentation preparation, and so on. The person costs should also be provided.

**Conservation Plan**

4.63 A conservation strategy should be outlined, if possible. It would include the number of tape rewinds (such as ten rewinds over ten years), the number of tape recopies (such as two recopies over ten years), and the amount of documentation requiring microfilming or other conservation treatment.

4.64 The cost of conserving the data file on an annual basis should be provided, if the information is available, as well as the person costs associated with the conservation plan.

**Servicing Plan**

4.65 The strategy which will be followed for servicing the data should be outlined. The most likely type of research request should be described (such as straight tape copy and documentation, extract, summary, statistical analysis, and so on). If any conversion of the archival format to a format suitable for research is required then this should also be identified. If possible, it would also be useful to describe the number of tapes involved in a typical research request and the number of potential requests per year. There should be mention of any special arrangements which have been made with the donor agency or some other agency concerning the servicing of the data. Any restrictions on access should also be identified.

4.66 And finally, the costs of servicing the data should be calculated, again if possible. This would include tape copying, documentation preparation, programming, and so on. The person costs should be included as well.

## SUMMARY

4.67 This chapter has described the procedures involved in the technical analysis of machine-readable records. Two technical factors have been mentioned which must be considered even before the content analysis is undertaken: whether the data (residing on either punched cards, magnetic tapes, or another storage medium) can be read; and whether there is sufficient documentation to appraise, process, and use the data. Other technical factors have also been described: the size of the machine-readable data file; the internal arrangement or structure of the data; the degree of software and hardware dependence; and the format in which the data will be preserved.

4.68 In order to assist in the application of the various technical factors, the second part of the chapter has provided a more detailed analysis of these attributes as they apply to specific kinds or types of machine-readable data. The latter have been restricted to survey files, time series files, and data bases. Each data type has been defined according to a set of commond technical characteristics, the major characteristic being the software environment within which each data type was originally managed. A set of question and answer planning tools has also been provided with each data type to permit a more systematic analysis of the technical consideraions.

4.69 It is only after both the Content Analysis and the Technical Analysis have been completed that a final evaluation decision can be made. In the last part of the chapter it is suggested that, if the evaluation decision is to acquire the data file, then it might be appropriate to develop a plan of action in terms of acquiring, processing, conserving, and servicing the data. Various factors to be taken into consideration have been described under each of these archival functions.

## NOTES

1. For a more detailed explanation of sampling techniques, etc., see Chapter V, paragraphs 5.24 to 5.33 inclusive.

2. Margaret L. Hedstrom, Archives and Manuscripts: Machine-Readable Records. Basic Manual Series, Society of American Archivists, Chicago, scheduled for publication in 1984.

3. For a more detailed explanation of the procedures involved in anonymizing machine-readable data, as well as the implications involved in undertaking anonymization, please refer to Chapter V, paragraphs 5.5 to 5.18 inclusive.

4. See John F. McDonald, The MRA and System 2000 Data Bases, a report of a special project undertaken in the Machine-Readable Archives Division of the Public Archives of Canada, April 1980.

## Chapter V

## ADDITIONAL ISSUES TO BE CONSIDERED

5.1 In the second chapter a number of factors were discussed which must be taken into consideration before machine-readable records can be appraised. They dealt essentially with the nature or manner in which machine-readable records are created and managed and the consequent problems which this poses to archivists when performing the appraisal function.

5.2 This chapter will address a number of issues which form part of the appraisal function and to which reference has already been made in those chapters dealing with the content and technical analyses of machine-readable records, more particularly questions associated with the personal nature of so much machine-readable data and the role of sampling in the appraisal process. The decision to treat these issues separately has been an arbitrary one essentially to permit a fuller explanation of some of the factors involved and some of the ways to proceed. Another issue that will be raised in the chapter is the implementation of a reappraisal policy and the value of such an approach for machine-readable records.

### APPRAISAL OF PERSONAL INFORMATION IN MACHINE-READABLE FORM

5.3 Many machine-readable records contain confidential information which documents the current status or recent past of many individuals, businesses, and organizations. Because archivists will give preference to the acquisition of microlevel data instead of aggregate data, and because they will make every effort to acquire copies of valuable machine-readable data files shortly after they are created, this will undoubtedly result in repositories holding large volumes of personally identifiable information of a contemporary nature. Added to this is the fact that some of the data files will be able to be linked to data from other sources. The creators of the machine-readable data files may be reluctant to transfer copies of such files to an archives because of the confidentiality of the information, the possibility of unauthorized access, and the ease with which the data can be copied.

5.4 Should an archivist be confronted with such a situation, it may be necessary to work out with the creating institution a procedure for the anonymization of highly sensitive data files. In such cases the repository will have two data files - one containing the raw or initial data, and the other containing the data in a public use or disclosure free format. It is not suggested that the anonymization

of data should be undertaken during the appraisal process. However, it may be necessary at this stage for the archivist to undertake a preliminary assessment of the value and results of anonymizing the data, as well as a guestimate of the person/time and computer costs which would be required. The creators and users of the data files should be able to provide some assistance with a number of these questions.

## Anonymization or Maintaining Confidentiality

5.5 Simply stated, the purpose of anonymization is to ensure that the unique identification of a single respondent or case in a data file is very difficult, if not impossible, to identify. In theory, a perfect anonymization process would require a wide range of inter-variable cross-checks to ensure that no combination of information in the file could be used to identify any one case. However, with data files of any significant size, such a process would be impractical, essentially because of the amount of work it would require. A more practical approach is possible, which is described briefly in the next few paragraphs, and which can also ensure the maintenance of a high degree of confidentiality.

5.6 Data files are generally composed of one or more of three broad types of information: demographic information, such as age, sex, marital status, and place of residence; socio-economic information, such as occupation, education, and income; as well as attitudes and opinions. Given the tendency for attitudes and opinions not to be unique to an individual, and the difficulty involved in finding a permanent record of an individual's attitudes or opinions at a given point in time, this type of data rarely leads to a unique identification of an individual in a file. In contrast, because much of the demographic and socio-economic types of information do exist in permanent records available to the public, it is these types that are the main focus of anonymization efforts.

5.7 In addition to the question of what types of information can be most easily used to identify cases in a file, which is primarily a question of the nature of the file's content, there is an aspect of size or quantity involved in maintaining confidentiality. Two factors are of importance. The first involves the number of response categories which a variable contains and the second involves the relationship between the number of respondents or cases in a file and the total population to which the file refers - the reference population. In general, the larger the number of unique response categories in a variable, the greater is the likelihood that one or more of the categories will contain very low frequency counts. Care must be taken to ensure that low response frequency categories

do not lead to the unique identification of cases in the file. As the frequency count in a variable's response categories rises, it becomes increasingly more difficult to identify specific individuals in any given category. As for the second aspect of size, the relationship between the number of cases in a file and the data's reference population, as the size of both the file and the reference population increases, case identification becomes more and more difficult.

5.8 As the purpose of anonymization is to prevent, as far as possible, the unique identification of cases in a file, a clear understanding is required of the ways in which identification can be made. Essentially there are three means by which cases can be identified: on the basis of a single unique characteristic; on the basis of a combination of characteristics that lead to specific identification; and on the basis of a combination of characteristics that lead to the identification of a limited number of qualified cases.

5.9 The most obvious targets of anonymization are the first type; that is, the unique characteristics. There are pieces of information or variables in the file that on their own uniquely identify the cases or respondents. Examples are data such as name and address, social insurance or social security number, drivers license number, and employee number. Characteristics such as these are unique to each case in a file and to each case in the reference population. In addition, information of this type usually already exists outside the context of the data file itself and therefore allows for accurate case identification. Name and address, for example, can usually be found in a telephone book or city directory with ease. All unique information in a file, on the other hand, may not necessarily lead to unique case identification. Often a file will contain a unique sequence number for each case, but unless some cross-reference between a case's sequence number and other identifying information such as name and address is available, sequence number alone will not identify a case.

5.10 The second way in which cases can be identified is on the basis of a combination of characteristics that lead to specific identification. While any one of the characteristics involved may not provide a unique identification of one or more cases, taken in combination they can lead to a unique identification.

5.11 The third means of identification involves a combination of characteristics that lead to the identification of a case in the file with a very limited number of respondents or cases in the reference population. Dealing with this type of possible identification is difficult because the appropriate action depends upon what is meant by the maintenance of confidentiality. If it means being

unable to identify a case in the file as being a particular individual in the reference population, then an adequate degree of confidentiality can be attained without a great deal of difficulty. However, if maintaining confidentiality means that information provided by a respondent will not be made available in a form that could possibly have an adverse and isolated effect on the respondent, then the criterion of unique identification may be too narrow. The difficulty involved in evaluating a file's sensitivity necessitates that, in the process of anonymizing a file, all information be treated as being of a sensitive nature. Where some of the information is clearly contentious, extra caution must be exercised to limit the possibility of even close, if not unique, identification of the cases in the file.

5.12 There are basically two approaches which can be taken with respect to the anonymization of machine-readable data. The first is by completely removing various pieces of information from the file; that is, deleting variables. The second approach is by aggregating the information contained in one or more variables in such a way as to eliminate the identifying characteristics by submerging it in successively broader categories of information.

5.13 The choice as to the proper or best approach to take with respect to anonymization, while based primarily upon considerations of maintaining confidentiality, should also involve an evaluation of the effect variable deletion and/or aggregation will have on the overall utility of the data file for further analysis. (1) In a sense this is a cost/benefit analysis of the process of anonymization - the benefit is the maintenance of confidentiality, while the cost is the resulting degree of information loss. Deleting or aggregating variables in a file always results in a reduction of the utility of the data for secondary analysis. Removing information from a file always represents an absolute information loss. Whatever amount of knowledge about the study group was contained in a deleted variable is completely unavailable to researchers, and since the information collected for inclusion in a data file generally has a specific, and important purpose, its elimination affects the usefulness of the other data. Similarly, aggregation represents information loss but usually not as completely as does deletion. The effect of aggregation is to reduce the degree of specificity of the data. As a variable is successively aggregated, fewer research questions can be addressed and fewer problems can be analysed. In brief, whichever approach is employed in anonymizing a file, a balance must be achieved between the need to maintain confidentiality and the effect of information loss on the utility of the data for research purposes. These are aspects which the archivist should address, at least in a general way, when undertaking the content analysis of the information. As mentioned previously, the level or degree of aggregation and the personal nature of the information can definitely affect the outcome of the appraisal.

5.14 As a means of maintaining confidentiality, deletion of one or more variables should always be a last resort. In general, this approach is required only in the case of information which assigns a single unique identifier to each respondent or case, such as the social insurance or social security number. As these types of variables are usually present in a file solely as a means of identifying each case rather than as pieces of information that are central to the research purpose of the file, they can be deleted without impairing the file's usefulness.

5.15 Where possible, aggregation should be the desired means to anonymize a data file. Essentially it is a process of recoding or combining the categories of a variable into successively higher or broader levels of generality in the factor being measured. Where the maintenance of confidentiality requires that aggregation be undertaken, there are basically two ways of recombining the categories of a variable. The first is to collapse or recode across all the categories. When this is done, all of the response categories of a variable are regrouped in such a way that the recoded variable measures the specific characteristic involved at a higher level of generality. While some information is lost in the process, such a recoding of the variable does maintain some of the original distinction of the variable. In general, this approach should only be used when the count in every category of a variable is small enough that the likelihood of being able to identify the cases involved is fairly high, either from the variable alone or when it is combined with other variables in the file.

5.16 A more common situation in anonymizing a file occurs when one or two categories of a variable contain only one or a small number of responses. Anonymization in such a situation can be undertaken in two ways. The first, and most desirable, is to combine the single category with another category that is very close conceptually. While this results in some information loss, it is restricted to the categories involved, and the detail in the other categories is maintained. However, in some situations there may not be another category of the variable that is conceptually close enough to the particular category requiring aggregation. In this kind of situation, it might be appropriate to combine a particular category with that variable's non-response category, thereby treating the single case as missing information on the particular variable. Recoding a variable's category to the missing data code is most often indicated when it is not a question of combining two conceptually close categories, but when there is not another category in which the problem category could be included. This is most likely to occur when the problem category is radically different from the other categories of the variable.

5.17 To be certain of which categories of which variables or combinations of variables could serve to identify one or more of the individual cases, it would be necessary to undertake a great deal of time-consuming cross-tabulation and evaluation. In files with many variables it is usually not practical to attempt so thorough an examination. While there are no rigid rules as to when to aggregate information, a general approach is that in any variable which provides very detailed information about a particularly "visible" characteristic of the cases involved, a single response category that contains three or fewer cases should be carefully examined, particularly where such a category represents a small portion of the reference population. This last qualification refers to the fact that while a variable category may have very few cases in it, if the number of people in the population from which the data were taken in that same category is large, identifying an individual may be very difficult.

5.18 Although almost any type of demographic or socio-economic variable could serve to identify particular cases in a file, the most common types of variables are those that relate to place of residence, occupation, education, physical possessions, and to a lesser extent variables such as age, year of immigration, religion, and ethnic background. These variables are of particular concern because the information they contain is often fairly readily available from sources other than data files. (2)

## Transborder Data Flow Implications

5.19 Transborder data flows is defined as "electronic or machine readable data or instructions which are transmitted or moved across national boundaries for purposes of processing, storage or retrieval in most cases utilizing computer-communication systems and interfaces." The convergence of computers and communications is developing as rapidly among countries as within each country. Indeed, the rapid rise of networks of interconnected computer systems and data transmissions on a world-wide basis has given traditionally domestic issues an international scope. In addition to the interconnection of national communication networks, the speed with which growing volumes of data are flowing across boundaries, and the "non-physical" nature of these data flows, has made many of the traditional methods whereby a country maitained its sovereignty either obsolete or irrelevant. The ability of the new technologies to collect, manipulate, store and transmit effectively and efficiently vast amounts of data have led to concerns related to vulnerability, dependency, and confidentiality, of which the protection of personal data (or privacy) is a most important aspect.

5.20 For practical purposes, physical national boundaries have little or no relevance to an increasing number of computer-communication networks. Because of this, the direct extension of the mechanisms that have maintained national boundaries in the physical world to those of the electronic digitized world may pose some problems. And yet this is the procedure which a number of countries have adopted. Such actions have usually been based on privacy concerns and consist essentially of "export" prohibitions on machine-readable data subject to certain conditions. It is mandatory that archivists be familiar with any legislation of this nature when assessing the legal value of machine-readable data.

5.21 The advent of electronically communicated transborder data flows has a large number of sovereignty-related consequences for nations. Fundamental among these is the inability to apply a country's laws to data stored outside the country. Needless to say, this has implications for the rights of privacy granted under a country's law to its citizens when personal information about the citizens is stored outside the country. It also has implications for maintaining the confidentiality of information, deemed to be confidential under a country's law, where the information leaves the jurisdiction. Other implications with respect to this issue are the potential frustration of government policies that require the production of information for decision-making purposes and court orders for the production of certain records, where the information and the records are located outside a country's territorial jurisdiction. At the same time, however, each country has full and complete jurisdiction to apply its laws to data originating in another country that are stored within its national boundaries.

5.22 The location of personal data banks abroad, in addition to privacy concerns, raises specific questions about cultural sovereignty. For example, if an international union were to automate its records-keeping activity and maintain its data bases and archives at its foreign headquarters, future researchers studying the labour history of a town or a certain industrial activity in the affected country would be obliged to go outside the country and request the permission of non-nationals to access the relevant information about their country assuming, of course, that the data still exist. Obviously, this is a cultural sovereignty aspect of transborder data flows which is quite unique and separate from any privacy considerations. (3)

5.23 Both nationally and internationally there is a great need for laws to be written or revised to meet the requirements of electronically communicated transborder data flows. It is certainly not the purpose of this study to explain what is required in this particular

field or how it should be done. By raising the subject of trans-border data flows, it has been the author's intention to raise a number of issues which may become of increasingly greater importance in the appraisal of machine-readable data.

## SAMPLING OF MACHINE-READABLE RECORDS

5.24 There may be occasions when the size of a machine-readable data file is such that the archivist may need to consider taking a sample of the information. While the sampling approach or criteria that the archivist uses may vary from one archival institution to another, as well as from one country to another, there are a number of basic principles which should be taken into consideration, regardless of the approach or criteria that are used. Therefore, this section will be restricted to a brief outline of those factors which should be taken into consideration if an archivist decides that a sample of the information should be taken as a result of his/her appraisal of the records. In many respects, the factors discussed are not unique to machine-readable records, but should apply equally to all information, regardless of the physical medium on which the information resides. (4)

5.25 Case files of government institutions usually contain information on all aspects of a country's population, its industry, and its development. They are large file blocks which essentially contain the same type of information. Their tremendous bulk often necessitates the implementation of sampling procedures in order to create a representative historical or archival record.

5.26 Concurrent with the creation of case files, government institutions create copies and extracts in both micrographic and machine-readable form. Indeed, more often than not, information in case files is in paper form, then replaced by microfilm to reduce bulk, and then changed to machine-readable form to improve programme delivery. It is not at all uncommon for a case file record on an individual or corporation to have a paper, microform, and EDP component at the same time. Therefore, if at all possible, the sampling strategy used by archival repositories should be developed in such a way that it is equally applicable to paper, microforms, and machine-readable records. The key consideration for sampling such records is the selection of those which are the most complete, not duplicated elsewhere, easily sampled, stored, serviced, and retrieved. In addition, for any sampling method to be viable, it is necessary to know at what rate, in what format, and by whom, case files are presently being created and what quantities are held in government institutions.

5.27 The archivist uses two general types of sampling: qualitative and quantitative. Qualitative sampling is the process as described in the third and fourth chapters of this study. In other words, under qualitative sampling or assessment, files or records are chosen on a qualitative or substantive basis, sometimes as much for artifactual and contextual interest as for their information or content value. Quantitative sampling or assessment, on the other hand, requires, or certainly should require, the application of statistical or scientific techniques.

5.28 For most government programmes, archivists will identify for archival preservation virtually all policy files, a significant number of operational files and, from time to time, a sample of case files. To date, most of the case file samples have not been taken on any scientific or statistical basis. Rather than being selected for representative information content, case files have often been retained to document the types of forms used, thereby providing an additional insight into the operations or administration of a particular programme. While this approach may satisfy traditional archival clientele, such as political and administrative historians, it is inappropriate for the changing nature of government records (that is, increasing volume, case file-based, and multi-storage technologies), as well as the changing patterns of social science research methods, activities, and interests (that is, "history from the bottom up," use of quantitative methods and techniques, and reconstituting the past by using linked microdata).

5.29 It is also important to keep in mind that, in the daily life of both business and government, sampling is a commonly used and highly regarded tool for administration, performance measurement, quality control, programme evaluation, planning, and so on. The last decade has also seen a rapid increase in the use of sampling and related quantitative methods by social scientists. This trend matches well with the fact that the case file has become the locus of interaction between the government and the individual, the corporation, and the organization.

5.30 A word of caution must be sounded, however. Too often sampling is done to reduce bulk and often fails to take account of the substance of the information in the records themselves. Many of the questions related to sampling involve the same considerations, analyses, and judgments that archivists must make when they are appraising the informational content of records. If the individual uniqueness or importance of the records is a prime consideration, such records should not be sampled. Therefore, sampling should be undertaken when the individual record or file is of no particular importance, but rather where the information in the record or file series will provide the researcher with an insight into a particu

lar aspect, change, or trend in society. In summary, then, sampling must not be a substitute for appraisal. It is merely a very powerful tool at the disposal of the archivist in implementing an appraisal decision.

5.31 There are a number of factors or parameters which might be considered by archival repositories as the sampling environment within which they might like to operate.

5.31.1 A sampling scheme for case files should be equally applicable to the paper, microfilm, machine-readable, and other storage media components and be appropriate to the various technologies utilized.

5.31.2 A sampling scheme should be cost-effective, easy and simple to implement, and administratively workable given the limited human and financial resources available.

5.31.3 A chosen sampling scheme should be simple and, if at all possible, uniform across all government agencies.

5.31.4 A sampling scheme should be consistent over a long period of time, maximizing the probabilities of creating longitudinal records and research possibilities.

5.31.5 A sampling scheme should maximize record linkage possibilities.

5.31.6 A sampling scheme for case files should minimize overall bias on a total or government-wide basis and, where bias is unavoidable for a particular sector, the degree of possible bias should be known.

5.31.7 A sampling scheme should readily lend itself to supplemental targetted sampling where deemed appropriate and necessary by the archivist.

5.32 The implementation of an archival sampling strategy which utilizes directed systematic sampling or any other sampling scheme, should enable a person to establish quickly and with a high degree of certainty whether or not a record on a particular individual, corporation, or organization exists in the archival repository. When sampling is undertaken, the entire case file series is destroyed except for the sample. It should not be possible to use such archival records for or against a company or person or third party because all other records of a similar nature have been intentionally destroyed in accordance with a records schedule. Furthermore,

these sample case file records would have been destroyed had it not been for a deliberate and conscious decision of the archivist to keep these particular records, not as individual records but as a group of records being representative of the characteristics of the case file series.

5.33 Therefore, it is extremely important that recorded information deposited in the archival repository as a result of sampling can no longer be used for administrative decisions which directly affect the party concerned, unless with the mutual consent of all parties involved. Otherwise, negative or positive inequities will result from the fact that certain information is still being available for some but not for others. Similarly, the archival sample should be considered to be legally "dead." In other words, recorded information existing as the archival sample should not be able to be used in any civil or criminal litigation or any enforcement of laws or regulations, unless all parties involved give their consent. This is because all other similar material is destroyed and the individual, corporation, or organization whose material is retained in the sample should not be able to use it to their advantage nor should such material be used to their disadvantage. (5)

## REAPPRAISAL OF MACHINE-READABLE RECORDS

5.34 Because of the substantial costs of long-term preservation, which includes the conversion of the data to current formats so as to prevent technological obsolesence, it is possible that not all machine-readable data files acquired by an archival repository will be retained forever. In order to determine which data files should be maintained and for how long, consideration might be given to the establishment of a reappraisal policy. (6)

5.35 As a practical way to implement such a policy, all machine-readable data files, upon acquisition by an archival repository, could be issued a review date. The latter would be suggested by the archivist and approved by the appropriate archival officer. The reappraisal or review date should be based on a number of considerations. One of these should be the length of time that the data could reasonably be expected to have value to the research community (such as five, ten, or more years). Another consideration might be the servicing conditions. For example, although the archival repository might have acquired a copy of a data file, the creating institution may have agreed to service the file until the end of the retention period. In such a situation the review date would be the retention period date, at which time the archival repository could choose to change the servicing conditions, the format of the

data, etc. The review date should also reflect any conditions or requirements of the transferring institution, any restrictions on access, or whether the data represent an only source or are part of a series or collection. However, no review date should be longer than thirty years and, for most data files, the review should be on a five to ten year basis.

5.36 Reappraisal should follow the same process as that of initial appraisal; that is, it should involve both a content and technical analysis, as described in the third and fourth chapters of this study. However, the process should not be nearly as time-consuming on this occasion, as the readability of the data and the availability of adequate documentation would already be known. More concrete information should be available at the time of reappraisal than may have been the case with the initial appraisal: the number and nature of research requests; the potential research activity; as well as the costs of storage and possible technical conversions. Where a number of files in a series or a collection of data files are to be reappraised at the same time, a representative sample need only be reviewed. Where one data file is being reappraised and is part of a series or collection, it should be reviewed in context with the series or collection, as its particular value might only be apparent as part of a larger whole.

5.37 All data should be assumed to be innocent until proven guilty. In other words, data under review should be judged to warrant continued archival retention unless the review clearly establishes that the costs, etc. of maintaining the data far outweigh their value for research.

5.38 There are a number of safeguards which should be taken when performing the reappraisal. First of all, the archivist should ensure that the file has had an adequate amount of time in circulation to serve public notice of its existence and availability. Secondly, consultation with current or potential users and researchers showed by considered. Thirdly, a panel of archivists might be convened in cases where the nature of the data and their archival value are in question.

5.39 If, as a result of the review process, it is clearly established that a data file should no longer be retained by the archival repository, then the file should be placed on probation for a period of time, possibly one or two years. During this probationary period, every effort should be made to ensure that the research community, other data archives and libraries, and particularly the creating institution, are informed that the data file will be eliminated unless its value can be re-established. Should an individual or

organization be interested in using the data, or should convincing arguments be made that the data should be retained, then the archival repository would undoubtedly retain the file and attach a new review date. In some cases, arrangements might be made to transfer the data file to the interested institution. On the other hand, should it be clearly established that no interest in the data has been expressed by any individual or organization, then the data will be eliminated at the end of the probation period.

5.40 Although the procedure as outlined above might be considered far too complicated and involved to justify the time required, it should be pointed out that, if the initial appraisal has been undertaken properly and thoroughly, the review process should result in a straightforward reconfirmation of the earlier decision. However, for those files that are reappraised as not having sufficient value to justify their long-term retention, it is the opinion of the author that the time invested in the reappraisal process is a worthwhile investment. This may be a minor cost considering the expenses involved in preserving machine-readable records over a long period of time.

## SUMMARY

5.41 This chapter has described in some detail the process of anonymization of data through either the deletion of variables or by means of aggregation. The advantages and disadvantages of each method has been explained, the more acceptable approach being aggregation. A number of issues associated with transborder data flows (that is, the transfer of machine-readable data across national boundaries as a result of communication and computer links) have been raised, with particular emphasis on personal information on one country's citizens located in a data base in another country. The second section of the chapter has dealt with the role of sampling in the appraisal process, with particular emphasis on the development of a sampling strategy that can be applied equally to textual records, as well as microforms and machine-readable records. The last part of the chapter has provided suggestions on how a reappraisal policy for machine-readable records might be implemented, and has described some of the factors which should be considered.

## NOTES

1. See Chapter III, paragraph 3.12.

2. A great deal of the information in this section on anonymization or maintaining confidentiality has been extracted from a Report on Data File Anonymization prepared by Kevin Selbee as a result of a personal service contract established with the Machine Readable Archives Division, Public Archives of Canada, August 1979.

3. The points raised in this section have been based on a Report on Sovereignty Aspects of Transborder Data Flows, prepared by Dr. J.V. Th. Knoppers in September 1982, under contract to the Canadian Department of Communications. This was one of several subjects which were addressed by an Interdepartmental Task Force on Transborder Data Flows, established in February 1981 under the coordination of the Department of Communications.

4. For a study of international sampling approaches or techniques, readers should consult Felix Hull, The Use of Sampling Techniques in the Retention of Records: A RAMP Study with Guidelines, General Information Program and UNISIST, UNESCO, Paris, 1981.

5. For a more detailed elaboration of the points raised in this section, as well as a full discussion of a proposed sampling strategy and methodology for the Public Archives of Canada, please refer to a Report on Archival Sampling Strategy and Related Issues, written by Dr. J.V. Th. Knoppers as a result of a contract with the Machine Readable Archives Division of the Public Archives of Canada, October 1983.

6. In his article, "No Grandfather Clause: Reappraising Accessioned Records," The American Archivist, Volume 44, Number 2, Spring 1981, pages 143-150, Leonard Rapport has applied the reappraisal proposal to textual government records. The suggestion for the establishment of a reappraisal policy for machine-readable records is based on Rapport's earlier proposal.

## Chapter VI

## GUIDELINES

6.1 This chapter is intended to provide readers with summary conclusions which are written in the form of recommended policies and practices, or guidelines. The numbers in parentheses refer to specific paragraphs of the study which contain a more detailed explanation of the subject(s) covered.

6.2 Archivists who are to be responsible for machine-readable records must become familiar with the basic terminology associated with data processing as well as with the operations of a computer system (1.2 to 1.41).

6.3 It is also important for archivists to be familiar with the nature of machine-readable records and how information in machine-readable form differs from other kinds of information, such as textual records and microforms (1.42 to 1.45). Machine-readable records have certain unique characteristics which must be known (1.46 to 1.49), as must the sources (1.50 and 1.51) and uses (1.52 to 1.55) of such records.

6.4 It is possible that some archival institutions may be unable to deal with machine-readable records because of limitations imposed by statutory or other regulatory authorities. This can include restrictions on the type or kind of records which the archival institution can acquire, as well as restrictions on the acquisition of recent records. There are a number of ways in which archival administrators can resolve these particular problems (2.3 to 2.9).

6.5 A number of issues arise when appraising machine-readable records with which archivists must be familiar. One is the existence of data in central government agencies which are often the compilation of data created by other government jurisdicitons, with no indication as to what governmental level owns the data and controls access to the data (2.11 to 2.14). A second issue is the control of machine-readable records that are created as a result of government contracts or research grants (2.15 to 2.23).

6.6 It is crucial for EDP records management programmes to be established in order for archival repositories to be assured of having a systematic acquisition programme for machine-readable data. In this way archivists can properly identify and appraise the machine-readable records that are created in the particular jurisdiction in which they work. While one of the major rationales used for a traditional records management programme has been the savings that can be achieved by storing voluminous quantities of records used infre-

quently in low-cost storage sites, cost-benefit analyses for EDP records management are still in the infancy stage (2.26 to 2.36).

6.7 In traditional records management policy and procedures, disposal plays a major role. However, such is not the case in the EDP world for, left on their own, those who control computer systems would automatically delete unwanted or unnecessary information. Because of this, it is imperative that records schedules for machine-readable information be established at the system design or planning stage for new applications or programmes (2.38 to 2.51). It is also important to remember that the archival limitations for information in machine-readable form may often be different from those for paper records (2.52).

6.8 Archivists will continue to work with records managers, at least to a certain extent, with respect to the scheduling of machine-readable records. However, it is the EDP personnel in the creating institutions with whom the archivists will need to work on a regular basis in order to ensure that machine-readable records are properly identified, inventoried, and scheduled.

6.9 The appraisal of machine-readable records involves the evaluation of the information contained in the records (content analysis), as well as an evaluation of the technical aspects of the records (technical analysis). The content analysis involves the traditional activities of archival appraisal combined with some new considerations particular to machine-readable records. Technical analysis, on the other hand, is a relatively new activity in the appraisal of records, but one which is of the utmost importance in the evaluation of machine-readable records.

6.10 Machine-readable records may have evidential value if they contribute to the policies or decisions adopted by a department or agency, or if they provide documentation of significant operations or procedures. Examples of machine-readable records which may have evidential value are provided in that section of the study which deals with the application of content analysis to individual categories of information (3.18 to 3.66).

6.11 Archivists must also consider the legal value of the machine-readable records which they are appraising. There are at least three factors which could affect the assessment for legal value. The first is whether or not such records are admissible as evidence in a court of law (3.5). The second factor is the association of the records with copyright law, both nationally and eventually internationally. Of particular importance are any special provisions to

cover computer programs or software (3.6). The third factor is the existence of any acts which stipulate that certain kinds of records must be retained for certain periods of time to meet particular legal requirements. This is especially the case when such acts include machine-readable data with their supporting documentation in the definition of "records" which must be retained for certain periods of time (3.7).

6.12 Another legal factor with which archivists must be familiar when appraising machine-readable records is the existence of any legislation which prevents the "export" of machine-readable records, usually containing personal information, from the country in which the records were created. This is how some countries have responded to the impact of electronically communicated transborder data flows. There are several other sovereignty-related issues associated with the transborder data flow question of which archivists should also be aware (5.19 to 5.23).

6.13 The main appraisal judgment in terms of content analysis is the value of the information the records contain for uses other than those for which they were created. The determination of informational value of machine-readable records is similar to the evaluation of other types of information for potential research value - an evaluation of the significance of the subject content for current and future research. However, there are a number of factors unique to machine-readable records which must be considered in appraising such records for their informational value. One of these is the uniqueness of the information or its format (1.46 to 1.49, and 3.9 to 3.11). A second factor is the potential for record linkage (3.13). Another important factor to consider is the level of aggregation (3.12 and 5.12 to 5.18).

6.14 The content analysis must be performed in consultation with departmental users, data processors, and other individuals connected with the information described in the file. It is important to keep in mind that the content analysis cannot be undertaken without the archivist having first obtained detailed information on the organization, the information structure of the organization, the purpose of the machine-readable data file, the methodology used, its use in the specific programme, its relationship to other programmes in the organization, and even its value in terms of the user's own perception of its worth to both the organization and potential research communities.

6.15 Before an assessment can be made on the historical or long-term research value of machine-readable records, it must first be determined if the information on the computer tape, punched cards, floppy disks, etc. can be read (4.4 to 4.6). It must also be determin-

ed if there is sufficient documentation accompanying the machine-readable records, consisting at least of a record layout and a codebook, to appraise and process the records and sufficient information for a researcher to use the records (4.7 and 4.8). If the data can be read and there is adequate documentation, then the archivist can proceed to the analysis of the contents of the machine-readable records and a more detailed technical analysis of the arrangement of the records and problems which could occur due to long-term storage.

6.16 In undertaking the detailed technical analysis, a number of factors must be taken into consideration. One of these is the size of the machine-readable data file. Should the size of the file pose difficulties, then the archivist might have to consider the possibility of obtaining only a sample of the records. In undertaking this, the archivist will have to determine the effect sampling might have on the informational value. It is important to keep in mind that sampling is not a substitute for appraisal. It is merely a very powerful tool at the disposal of the archivist in implementing an appraisal dicision (5.24 to 5.33).

6.17 Another factor which must be addressed when undertaking the detailed technical analysis is the internal arrangement of the data. The arrangement of the individual records on the reel of tape is rarely a major consideration, but the character codes used and the dependence on certain computer programs could have a major impact on the processing of the data (4.11 and 4.12).

6.18 The major onsideration when undertaking the technical analysis is the hardware dependency of various storage media and the software dependency of certain formats of information. In both cases archivists must be aware of the costs associated with reformatting the data should this be required (4.13 to 4.16).

6.19 It must also be remembered that any machine-readable records which are acquired must also be preserved. During the technical analysis the archivist should determine, if at all possible, the costs which will be required to preserve the data for a long period of time (4.17).

6.20 Two additional factors should also be considered. The archivist will need to determine who will fill service requests on the data - whether the originating department or the archival repository. Should the data file be software and hardware dependent, it might be decided for the originating institution to handle all service requests. The nature of any restrictions on the data must also be determined. While the same kind of restrictions apply to machine-readable records as to textual records, the manner in which such

restrictions are handled is different. If only certain portions of information are restricted, it is possible to remove all other portions of information from the file for research use, thereby creating a public use version of a restricted file. However, it is important for the archivist to consider the impact such measures would have on the informational value, as well as the cost of producing a public use version (4.18 and 4.19, as well as 5.5 to 5.18).

6.21 The analysis of the technical consideraions of a machine-readable data file should lead to a more rational development of an approach to the acquisition, processing, conservation, and servicing of the data. The approach itself should be developed according to the willingness of an archival repository to absorb the costs associated with each of these archival functions. In order to assist in the evaluation of such technical attributes as software, hardware, size, and physical arrangement, and in order to provide a more systematic analysis of the archival functions, archivists may wish to use question-and-answer planning tools which can be developed for different kinds or types of data (4.35, 4.40, and 4.53).

6.22 It is at this stage that the archivist should being together the results of the content analysis and technical analysis and justify the decision to acquire the records or to reject the records. Should the appraisal decision be favourable, it is suggested that a plan of action be developed, using information contained in the planning tools referred to in paragraph 6.21 above. Such an action plan could cover the acquisition, processing, conservation, and servicing functions (4.58 to 4.66).

6.23 It is possible that, because of the substantial costs of long-term preservation which includes the conversion of the data to current formats so as to prevent technological obsolescence, not all machine-readable data files acquired by an archival repository will be retained forever. In order to determine which data files should be maintained and for how long, archival administrators might wish to consider the establishment of a reappraisal policy. As a practical way to implement such a policy, upon acquisition by an archival repository all machine-readable data files could be issued a review date (5.35 to 5.40).

# APPENDICES

# APPENDIX A

# APPRAISAL FORMS AND PROCEDURES

## A.1

## SELECTION CRITERIA

## DEVELOPED BY THE

## PUBLIC RECORD OFFICE

## AND AN

## ASSESSMENT

## OF

## DEPARTMENTAL ADP FILES

## MACHINE READABLE RECORDS FOR PERMANENT PRESERVATION

### 1. Selection

1.1 In appraising computer files the PRO has formulated criteria for selection which do not conflict with the recommendation of the Grigg Committee "that particular instance papers should be examined to determine what papers, if retained, would give the greatest amount of information in the smallest amount of space."

The criteria adopted are:

a) the information content should be high;

b) the information available should not be available from conventional records in an acceptable form;

c) the record should be a main data set not a compilation from various sources available in this or another media elsewhere;

d) the information should not be aggregated to any significant degree;

e) the information should be susceptible to analysis for purposes beyond that for which it was collected; and

f) small data sets only to be taken where there are positive advantages in taking data in machine readable form in preference to the conventional records from which they were compiled.

1.2 These criteria, amplified by normal selection criteria have led to the identification of the files listed in Annex A.

1.3 Annex B of this paper lists the main task undertaken by computer by a number of departments.

### 2. Practical Considerations

2.1 In the course of its work on machine readable records the PRO has found a variety of problems: the most significant are those of time scale, changing media and format and cost.

2.2 The time scale laid down in Section 3(4) of the Public Record Act 1958, that is that selected records must be transferred when they are 30 years old, is inappropriate for machine readable records. In particular:

a) documentation cannot be corrected so long after the records were created; and

b) the impermanence of data on tape or disc make it essential that proper maintenance is carried out on selected material to ensure survival.

2.3 Some of the files identified as potential for permanent preservation have already changed their format and many may now he held on disc rather than tape. Some of these changes are expensive in staff and machine resources.

2.4 The cost of maintaining tape has been estimated as between £30 and £70 per tape per year with the higher estimate thought reliable. It will be appreciated that the upkeep of a modest archive of say 5,000 tapes where the majority will not be available for public use for another 25 years is a major financial burden, e.g. the cost of the above would be £8.75 million over 25 years.

With these kinds of expenditures involved we need to be as sure as possible that whatever is selected is likely to be of real use and be used.

In a list of priorities how high on the list should be machine readable records? Should the amount of conventional records retained be reduced?

3. Conclusion

Machine readable records have a place in our archives. The cost may appear high at the moment but technical changes will probably lower these costs. We cannot ignore them; they are now getting too numerous. We must be aware of what is available and very selective in our appraisal for permanent preservation. We should also be sure that tapes no longer required are not retained occupying valuable space and possibly computer time.

D Ashton

## DEPARTMENTAL ADP FILES

1. **AGRICULTURE, FISHERIES AND FOOD**
(Computer ICL 1904S)
No files earmarked for preservation, but the Agricultural Census would be accepted if an amendment to the Agriculture Act 1947 allowed disclosure of information.
Potential input 8 tapes per year.

2. **CUSTOMS AND EXCISE**
(Computer ICL 4/72 & 2966(2))

2.1 **VAT**
Summary Trader Information.
Input 7 tapes per year.

2.2 **IMPORT/EXPORT**
The import/export statistics files are a possible item for transfer.
Potential input 48 tapes per year.

3. **DEFENCE**
(Computers various)
Nothing worthy of permanent preservation has yet come to light, but our enquiries have not been as thorough as possible owing to limited resources.
Potential input not known.

4. **EDUCATION AND SCIENCE**
(Computer ICL 1906A)

4.1 Immigrant Pupils 1969-1973 Survey.
Received 4 tapes.

4.2 Potential
There is potential for taking further teacher and pupil statistics.
No files defined.
Potential not known.

5. **EMPLOYMENT**
(Computer ICL 1904S)
Nothing yet transferred.
Family expenditure survey selected for preservation (Files back to 1961).
Input 4 tapes per year.

Potential
Further files could be taken in with an amendment to the Statistics of Trade Act 1947 e.g. Labour Cost Survey, Census of Employment Characteristics of Unemployed.
Potential 30 tapes per year.

6. **ENVIRONMENT**

(Computer PRIME 550 ICL 1904S)

Files on the National House Condition Surveys and Local Authority and Building Society Mortgages have been received.
Received 4 tapes.
No further files have been nominated but there is some scope.
Potential 4 tapes per year.

7. **FORESTRY COMMISSION**

(Computer IBM 1130)

Files on the Tree Census and Dutch Elm Disease Survey received.
Received 100,000 cards (now on magnetic tape).
Other Surveys on small mammals and birds are probably worth keeping.
Potential 1 tape per year.

8. **HEALTH & SOCIAL SECURITY**

(Computers ICL 2980 & 2956)

The following annual statistical files have been selected for preservation:

1) Supplementary Benefit
2) Pneumoconiosis Bysinosis and Miscellaneous Diseases
3) Sickness Injury Benefit
4) Disablement Benefit
5) Prescribed Diseases
6) Industrial Death Benefit
7) Unemployment Benefit
8) Insured Population

Approximately 500 tapes are outstanding plus 8 per year accumulating. We have agreed to take a copy of the "Newcastle Records Main File" when the time is appropriate. Records on Mental Health are also to be transferred when possible.
Potential 220 tapes every 10 years.
Records at the Reading Computer Centre have been considered but only the records concerning drugs, which have some confidentiality are considered possible.
Potential 1 tape per year.

9. **HOME OFFICE**

(Computer ICL 2960/10(2))

The following files ahve been earmarked for preservation:
National Criminal Statistics 1970-1973
National Criminal Statistics 1974-
Prison Parole Index
Prison Index
Race Relation Board Statistics
Immigration Board Appeal Statistics
(Due to the confidentiality of these files and consideration of the data protection aspect further review has been halted.)
Potential input 80 tapes per year.

10. **INLAND REVENUE**
(Computer IBM 3081)
The following statistics files have been earmarked for preservation:
Income Survey
Schedule "D" Assessments
Corporation Tax
Capital Transfer
Potential input 4 tapes per year.

11. **METROPOLITAN POLICE**
(Computer ICL 1904S)
One file on "Blood Groups" earmarked for preservation some time in the future.
Potential input 1 tape.

12. **OFFICE OF POPULATION CENSUSES & SURVEYS**
(Computer ICL 1960/10)
We have agreed to take magnetic tapes for the Censuses from 1961; but as these records are still required by OPCS and being maintained by them in a responsible manner we do not envisage taking these files into the PRO within the next five years.

Potential input
1961 Census 100% 42 reels
1961 Census 10% 13 reels
1971 Census 10% 30 reels
1971 Census 100% 200 reels
1981 Census 100% 200 reels

There are also a large number of surveys taken each year and the General Household Surveys has been earmarked for preservation. Social Survey Division are now considering their attitude and policy towards preservation of other surveys.
Potential 2 tapes per year.

There is additionally a number of tapes on population which the department are considering if they can be transferred to PRO but are felt to be covered by Section 4.2 of the Population Statistics Act 1938.

13. **TRADE & INDUSTRY (Including BSO)**
(Computer ICL 2982)
The Statistics of Trade Act 1947 bars disclosure of information but an amendment to this Act could mean a considerable input into the PRO.
Potential 20 tapes per year.(?)

14. The following departments have computer files which are at present not thought worthy of preservation:
Central Statistical Office
Energy
Land Registry
Lord Chancellor's Department
Meteorological Office (preservation on site)
National Coal Board
Ordnance Survey
Post Office.

15. In addition to the above a number of ad hoc surveys are undertaken and a number of these each year might merit preservation.
Potential 4 tapes per year.

## MAIN TASKS ON COMPUTER: GOVERNMENT DEPARTMENTS

1. **Agriculture, Fisheries & Food**

1.1 Departmental book-keeping, vote accounting and budgetary control.

1.2 Guarantee payment systems for fatstock.

1.3 Payment of grants and subsidies including EEC schemes.

1.4 Monthly and weekly payrolls and personnel records.

1.5 Agricultural censuses and economic surveys including Farm Management Survey.

1.6 National Food Survey.

1.7 Farm Wages and Employment.

1.8 Agricultural Rent Enquiry.

1.9 Slaughterhouse statistics.

1.10 Preparation of accounts and statistics for cattle breeding centre.

1.11 ADAS scientific and technical projects, e.g. dairy management, dairy herd long-term forecasting, soil analysis, etc.

1.12 Fisheries research including analysis of biological, radio-biological and hydrographic data.

1.13 Fisheries catch statistics.

1.14 Standard statistical techniques.

1.15 Charting progress of brucellosis eradication.

1.16 Centralized addressing.

1.17 Agency work for Intervention Board for Agricultural Produce.

1.18 Dairy Management Scheme - Gross Margins.

1.19 EEC Farm Structure Survey.

2. **Customs & Excise**

2.1 VAT: Maintenance of the file of Registered Traders; issue of Tax Returns, Reminders and Assessments; accounting for payment of tax and the issue of repayments; VAT control work; generation of management, trade and revenue statistics; Batch and on-line interrogation facilities.

2.2 Trade and Revenue Statistics of UK; collation of data, preparation of monthly and Annual Accounts and production of special tabulations to other Government departments and EEC.

2.3 Computerised control of stocks in Bonded Warehouses and preparation of Revenue Statistics relevant thereto.

2.4 Preparation of daily and periodic accounts of monies collected in offices of Customs and Excise throughout the UK.

2.5 Control of Duty Deferment scheme including supply of magnetic tapes containing relevant direct debit details to Bankers Automated Clearing Services Ltd.

2.6 Staff transfers and postings.

2.7 Departmental Costs and Resource Allocation Analyses.

2.8 Central Filing Office record.

3. **Defence**

3.1 Stores Control.

3.2 Personnel Records and Pay.

3.3 Ships Design.

3.4 Dockyard.

3.5 Various scientific applications, e.g. Wind tunnel, flight systems, Laboratory data analysis and Blind landing research.

4. **Education and Science**

4.1 Maintenance of the Main Mechanized Record of Teachers (MMRT) and the production of Directories of Teachers.

4.2 Processing of data mainly within the MMRT, to aid the administration of the Teachers' Superannuation Regulations.

4.3 Production of Teachers' Statistics for internal Departmental use, for publication, for use in Government Actuary's valuation and for the Burnham Committee.

4.4 Production of Schools Statistics for Departmental use and publication - annual census of pupils and staff in English maintained primary, middle, secondary, nursery and special schools, also English and Welsh direct grant and independent schools.

4.5 Annual sample survey covering leavers from maintained and independent schools in England and Wales, which provides analyses of numbers of leavers by age, qualifications and destination. The participating schools also provide census returns of candidates to GCE/-CSE examinations. The statistics are used for Departmental and publication purposes.

4.6 Sample survey of GCE "A" level achievements in further education establishments in England and Wales. The establishments also provide census returns of candidates to GCE/CSE examinations. The statistics are used for Departmental and publication purposes.

4.7 Survey of GCE/CSE Examination Board candidate entries and results. The statistics are used for Departmental and publication purposes.

4.8 Production of statistics from sample surveys of selected schools for HM Inspectorate.

4.9 Index of Educational Establishments - maintenance of a master index file of schools, further education and other educational establishments.

4.10 School Buildings - maintenance of file of school building projects.

4.11 Maintenance of DES Vote and Non-Vote Accounts.

4.12 Voluntary Schools - maintenance of Grant payment records for school building work.

4.13 Censuses of libraries in major establishments of Further Education and staff in librarianship and information service work.

4.14 Production of statistics of students enrolled on Courses in Colleges of Further Education (including University Departments of Education). Student information is collected via a Data Interchange System.

4.15 Survey of first destination of polytechnic students.

4.16 Production of Teacher Service Cards and Status Letters from records of students in final year of Teacher Training.

4.17 Adult Education Survey.

4.18 Post Graduate Student Award System.

5. **Employment**

5.1 Payroll for DE and the Agencies.

5.2 Statistics, including New Earnings Survey, Family Expenditure Survey, Census of Employment, Retail Prices Index and Unemployment Statistics.

5.3 Manual Giro Reconciliation.

5.4 Training Opportunities Scheme Statistics.

5.5 Bureau service for the DE Group.

5.6 Medical Surveys for HSE.

5.7 Professional and Executive Vacancy matching system.

5.8 Management Information for the Employment Service and Training Services Division of the MSC.

6. **Health & Social Security**

6.1 Main Tasks.

6.1.1 The primary tasks undertaken are the recording of all contributions paid under the Social Security Acts and the provision of an enquiry service for DHSS and DE local offices. The main contribution record file contains about 46 million accounts; this is held on magnetic tape and is searched daily.

6.1.2 Another system, reconciles over 100 million girocheques issued each year by DHSS and DE.

6.1.3 Retirement and Widows Pensions.

Payment records of some 7½ million pensions are held on magnetic tape and are processed to issue order books (printed on MDS off-line printers and made up by CEM order book assemply equipment), payable order and girocheques. In addition the records of 1½ million cases where Retirement Pensions are paid together with Supplementary Bene-

fit by Local Offices are held on the main file in order to provide the Local Offices with an updating service in respect of the Retirement Pensions. The complete file is updated daily.

6.1.4 Child Benefits.

The computer system maintains records and issues payments for over 7 million families entitled to this allowance. The main file is updated daily and 18 week order books are computer produced, printed on MDS off-line printers and assembled on CEM machies in a similar way to the Pensions order books at Newcastle.

6.1.5 Payroll.

The Payroll system covers some 94,000 staff of the Department and maintains personnel information.

6.1.6 Statistics.

6.2 Reading.

6.2.1 Over 96% of computer processing carried out at Reading is in support of the National Unemployment Benefit System. The installation is linked by data transmission to Department of Employment Unemployment Benefit Offices (UBOs).

6.2.2 In March 1980, some 296,000 girocheques, with a cash value of about 13.8 million, were being issued weekly, and the system was handling about 600,000 claims for benefit, representing 45% of the National load. The system has a design load of 800,000 claims, with contingency provision for an additional 100,000 claims.

6.2.3 Other computer applications undertaken at Reading include:

6.2.3.1 Drugs Licensing - this system maintains pharmaceutical and other data for licensed medicinal products in a form suitable for administrative purposes and interrogation using FIND-2 and FILETAB.

6.2.3.2 Drugs Monitoring - maintains non-personalised patient records of suspected adverse reactions to drug therapy, as well as certain laboratory data, in a form suitable for interrogation by a suite of programs written in BCPL, PLAN, COBOL, and FORTRAN. A large number of standard tabulations are also provided at regular intervals.

7. Ordnance Survey

7.1 Sales invoicing, credit control and sales analyses, by means of new in-house system implemented in April 1978.

u

7.2 Control of, and accounting for fixed assets, and production of a vocabulary of stores.

7.3 Production of digital maps both large scale and small scale derived from larger scale data by means of a much enhanced replacement system introduced in late 1977.

7.4 Calculations for air and ground surveys and levelling.

7.5 Management statistics. Unified management and accounting information is provided for 4 Divisions by means of the ICL NIMMS package.

7.6 Long-term planning model.

7.7 Scientific survey network adjustments and analysis.

7.8 The generation and typesetting of marginal text on maps.

7.9 Processing of paper tape output from automatic recording planimeters used for measuring areas; a by-product of the calculations in a paper tape for film setting.

7.10 Typesetting for gazetteers and for map sheet edges.

8. **Office of Population Censuses and Surveys**

8.1 Censuses of population.

8.2 Production of statistics from records of births, deaths and various other population and medical data.

8.3 Ad hoc social surveys.

8.4 General Household Survey.

8.5 Labour Force Survey.

8.6 Preparation of quarterly indexes of all births, deaths and marriages registered in England and Wales.

# APPENDIX A

# APPRAISAL FORMS AND PROCEDURES

A.2

EXTRACT

FROM THE

"DISPOSITION OF FEDERAL RECORDS,"

A RECORDS MANAGEMENT HANDBOOK

DEVELOPED BY THE

NATIONAL ARCHIVES AND RECORDS SERVICE,

GENERAL SERVICES ADMINISTRATION

(7610-01-055-8704)

# III. THE SCHEDULE: RECORDS VALUES AND THE APPRAISAL PROCESS

- Appraisal is the evaluation of records to determine whether they are permanent or temporary.
- It is a joint responsibility of the agency and NARS.
- Records values are
  Primary: of interest to the agency.
  Secondary: of interest to future researchers and the public.
- Both values are equally important.
- Most permanent records reflect the continuity of Government.

**Appraisal Objectives**

After the inventory has been completed, the value of each series for the agency and for others must be determined. This process is called appraisal. Appraisal places all series into one of the following categories:

- Records of permanent value, to be preserved by NARS.
- Records that are disposable, immediately or later.

Disposition specialists play an important role in the appraisal process. Although NARS has final responsibility for appraisal judgments, agency specialists should be familiar with appraisal objectives and standards. Agency specialists use those standards in drafting schedules.

The determination of records values is basic to records disposition and archival management. Techniques cannot be devised that will reduce to a mechanical operation the work of deciding on values. Success depends on

- Understanding how agency activities are documented.
- Appreciating the function of archives and the importance of our understanding of the past.

Above all, appraisers must realize that some records values go beyond the immediate interests of the agency. Agency needs are important. Equally important are the data available in records long after they have ceased to have any value to the agency.

**Inventory Analysis**

As a first step in the appraisal process, series inventory forms should be examined to get a broad view of how well agency functions are documented. The appraiser should always be aware of three characteristics of Federal records common to all agencies, as illustrated in figure 11:

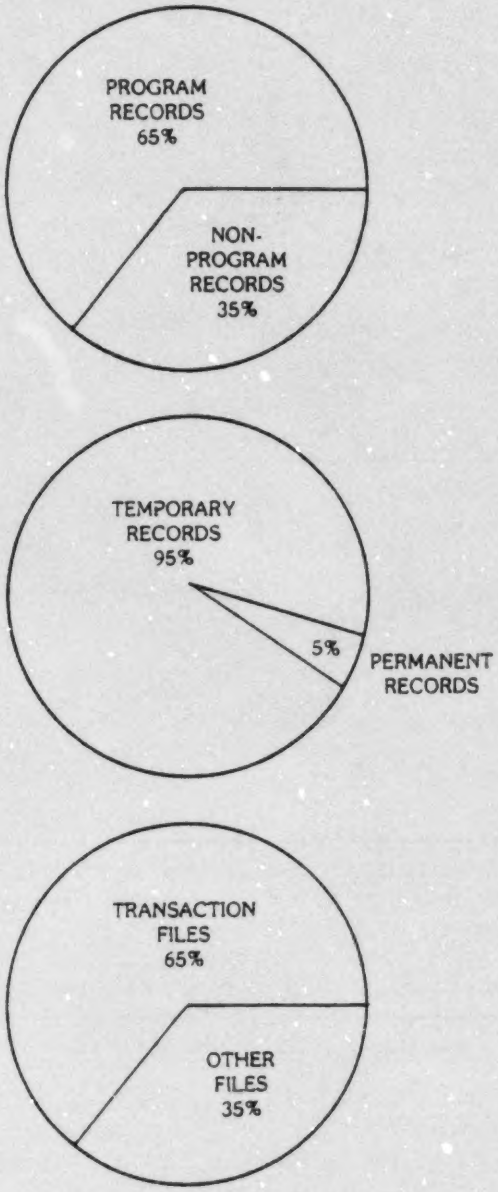



**Figure 11.—The Appraisal Process: Three Aspects of Federal Records**

- The ratio between the volume of program records and nonprogram, or housekeeping, records.
- The ratio between expected volumes of permanent and temporary records.
- The ratio of transaction files to all other records.

The appraiser must ask:

- Which records seem to reflect policymaking?
- Are there summary records? Where?
- Does the inventory distinguish between records produced by
  a. Housekeeping, or routine administrative, activities?
  b. Program, or important agency, activities?

**Values: Primary and Secondary**

Based on long experience with administration in agencies and with historical research in NARS, it is evident that records values cannot be reduced to exact standards but consist of little more than general principles, or yardsticks. All records values fall into two broad categories:

**Primary Values:**
**The Agency Point of View**

These values are of interest only to the agency, which keeps records as long as they are needed for administrative, legal, and fiscal uses. Once the records are not needed for current operations, a disposition program provides for their removal from office space so that their sheer bulk does not impede administration. Agency appraisers are responsible for judging the usefulness of records for current or future operations of the agency.

**Secondary Values:**
**The Archival Point of View**

These values exist after records are no longer useful to agency officials.

---

TABLE 3 - VALUES OF RECORDS:
AN OVERVIEW

| | |
|---|---|
| **PRIMARY VALUES** | |
| Administrative | |
| Legal | |
| Fiscal | **AGENCY INTERESTS** |
| Scientific | |
| **SECONDARY VALUES** | |
| Evidential Values | |
| Data on Agency Origins | **SHORT-TERM INTERESTS:** |
| Data on Agency Programs | Agency |
| Data on Agency Policies | **LONG-TERM INTERESTS:** |
| Data on Agency Procedures | |
| | Other Agencies |
| | Future Researchers |
| Informational Values | |
| Data on Persons | |
| Data on Things | **LONG-TERM INTERESTS:** |
| Data on Places | Future Researchers |
| Data on Phenomena | |

---

Their interest to others is based on the long-range needs of research and scholarship. Valuable records are preserved by NARS so that the public can use them.

Although NARS has prime responsibility for determining the secondary values of records, an agency can assist in this evaluation. The archivist contributes an understanding of research, and the agency specialist contributes important information about the creation and use of the records.

**Aspects of Primary Values**

Following is a discussion of primary values. Many series may have more than a single primary value.

**Administrative Value**

Records have this value if they help an agency perform its current work. The time during which this value exists may be long or short, depending on the purpose it serves. Records such as routine requisitions have a short-term value. Closed transaction files may pertain to long-term fiscal, regulatory, or control operations, and their administrative value may last over a long period. Directives, orders, and regulations obviously have long-term administrative value.

Many records at operating levels have little administrative value because they are

- **Duplicated elsewhere,** as in correspondence or directives.
- **Summarized at higher agency levels,** as in reports, raw data, and working papers.
- **Temporary controls,** such as logs and tickler files.

Most housekeeping documents have short-term administrative value because they document routine transactions quickly completed. They may consist of requisitions, purchase orders, stock control records, personnel records, and the like.

**Legal Value**

Records have this value if they contain evidence of legally enforceable rights or obligations of the Government. Among those obligations are the legal rights of persons to make claims against the Government.

Among records having legal values are

- Legal decisions and opinions.
- Documents involving legal agreements, such as leases, titles, and contracts.
- Evidence of actions in particular cases, such as claims papers and legal dockets.

The duration of legal values varies with the matter at hand. For example, legal values to the Government of contracts and claims records diminish rapidly after final settlement of the contract or claim. It may cease upon expiration of any pertinent statutes of limitations.

**Fiscal Value**

Records having fiscal value relate to all financial transactions. They may be budget records, which show how expenditures were planned. They may be voucher or expenditure files of several kinds, which document the purposes for which agency funds were spent, or they may be accounting records. The form and content of many fiscal records are prescribed by various staff agencies: Office of Management and Budget, General Accounting Office, Treasury Department, General Services Administration, and others. In most instances, only the data on the forms differ from agency to agency. Records relating to the development of fiscal policy should not be confused with those of fiscal transactions. Fiscal policy files may have permanent value.

**Scientific and Technological Value**

Records having scientific and technological value consist of large quantities of technical data gathered as a result of pure and applied research. If the data are not used immediately or if the research results remain unpublished, the records may then require a lengthy retention period.

The nature of scientific inquiry may make appraisal difficult. Series long dormant may suddenly become a link in a chain resulting in a new discovery. While these circumstances are not easy to predict, agency scientists can provide some guidance.

**Aspects of Secondary Value**

Secondary values, which go beyond agency needs and interests, are of two types:

- **Evidential values.** These relate to evidence the records contain of the organization and functioning of the agency that produced them.
- **Informational values.** These relate to information the records contain on persons, things, problems, and conditions with which the agency dealt.

The legal definition of records discussed in chapter I shows the distinction between evidential values and informational values. As mentioned previously, the word "records" is partially defined in the Records Disposal Act of 1943 to include all materials preserved "as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government. ..." These kinds of records relate to agency origin,

development, and accomplishment and are known as evidential records. They contain the evidence of agency achievement. The statute further defines "records" to include materials that may be preserved "because of the informational value of data in them. ..."

Although the majority of records created by the Federal Government are textual, records are not restricted to correspondence, memorandums, reports, and similar documents prepared on paper. Official Federal records also include microforms, photographs, motion pictures, sound and video recordings, maps, architectural drawings, aerial photographs, and machine-readable tapes and discs. Records may have secondary values regardless of their form. Therefore, decisions concerning which records to preserve and which to destroy relate mainly to the content of the record, not its form.

**Evidential Values**

Evidential means evidence of agency organizational structure and functions. Some records containing such evidence have archival value. Such records are needed as evidence of agency responsibilities and how it discharged them. They are neded so that the experiences of an agency can be examined later.

Records should be preserved that contain evidence of an agency's "organization, functions, policies, decisions, procedures, operations, or other activities." These records should show an agency's origins, its administrative development, and its organizational structure. They should show the policies it followed, the reasons for their adoption, and the procedures used to implement the policies.

Evidential records provide precedents so that when quick decisions are needed much planning effort can be avoided. For students of public administration, historians, and other social scientists, the records are sources for evidence of how the Government conceived of the needs of its citizens and how it served them.

**What Records are Evidentiary?**

- **Records relating to agency origins** reflect the conditions that led to its creation:
  a. Statutes.
  b. Executive orders.
  c. Investigations by Congress or other bodies.
  d. Other scattered documentation.
- **Records relating to agency policy and procedure** must be singled out for special attention. Such records reflect the communication of policies and procedures to offices.

In general, the records kept should show agency organization, plans, methods, and techniques used to carry on its business. The ease in locating policy and procedure materials depends on the quality of agency documentation. Quality, in turn, depends largely on how well decisions are recorded and on whether or not the records are filed separately from temporary materials.

Sources of policy and procedure records are the following:

a. **Directives, manuals, and handbooks**, together with related endorsements, comments, and clearances. Sets of all issuances (current as well as obsolete) should be kept.

b. **Organization charts and directories** provide important documentation of the structure of an agency. Master sets should be earmarked for preservation. Sets of telephone directories should be retained. The organizational part of the directories provides data on organizational structure.

c. **Narrative and statistical reports** on accomplishments at divisional or higher organizational levels are important. Often there exist summary narrative accounts of the direction and execution of agency programs, which have been prepared for annual reports or for other purposes. Narrative accounts of the history of an agency, produced by official historians, may have an even greater value than annual or other periodic reports of accomplishments and should be retained.

d. **Publicity material**, in the form of press releases, official speeches, charts, and posters, showing the actual administration of programs.

Other policy records are not easily found or identified. Digging them out of agency records requires careful analysis. Records of top management, whether located in offices or central files, should be retained. The records of key officials may include their correspondence files, minutes of conferences and staff meetings, memorandums, directives, and other evidence of official action. (See table 4.) In general, the records of offices decrease in value as the administrative ladder goes down. The extent to which an appraiser goes down that ladder to get important policy documentation varies with each agency. It depends on organization, delegations of authority within the agency, the reporting system, the filing system, and the documentation system. The appraiser must find and identify the policy records and, if possible, fill in gaps from records at lower operating levels. In the process, the appraiser will find that most records at operating, or nonexecutive levels, do not reflect policy, but instead:

a. Concern housekeeping matters.

b. Are temporary aids to make program operations run more smoothly.

c. Are transaction case files.

The first two kinds of records usually are given short retention periods. Although transaction files as a whole rarely have permanent value, selected samples may be retained.

**Informational Values**

Informational values relate to the information that is created as a result of agency programs. Evidentiary values, discussed in this chapter, deal with the agencies themselves. In looking for informational values, the appraiser is not concerned with the agency that created the records or what programs they involved. The concern, instead, is with the information in them. The information relates to persons, things, places, and phenomena and is as diverse as the work of the Federal Government. Several considerations must be carefully weighed when records are appraised for their informational values. First, records with proven or anticipated reference value are normally designated for transfer to the NARS. Second, unnecessary duplication of information should be avoided by comparing potentially permanent records with similar records created at different organizational levels. Generally, only the most complete or "record" set of documents is designated for preservation. In isolated cases, two sets of records with the same historical information may be identified for permanent retention if they are arranged differently (for example, chronologically as opposed to alphabetically by subject) and their different arrangement patterns facilitate reference use. Third, the significance of programs varies from agency to agency and consequently, appraisals of records must also vary. Hence, the number of series of records selected for permanent retention from each agency will depend on the importance of its programs.

There are three tests for informational values:

- **Uniqueness.** This means that the information in a group of records cannot be found elsewhere in as complete and usable a form. "Elsewhere" can mean Federal or non-Federal sources. In these cases, the appraiser must know or consult archivists who know about other possible sources of data on a given subject.

- **Form.** This test mainly concerns the degree to which the information is concentrated. It relates also to the physical condition of the records. Ease of access to the data often needs to be considered when choices must be made between two series, arranged differently, containing virtually the same documentation. Choices may have to be made between paper records, machine-readable records, or records stored or processed by other media, but data concentration and ease of access are still the important tests.

- **Importance.** The research importance of a series is an educated guess by an appraiser. This concern should involve consultation with the NARS staff. Such decisions must be based on a broad knowledge of re-

search techniques and the scholarly world. Familiarity with American history and with the interests of social scientists and the general public is also important. This knowledge is brought to the appraisal process by NARS specialists.

Information values concern

- **Persons and corporate bodies**, such as boards, commissions, and firms.

The problem of deciding which records to keep concerning human beings is a difficult one. Most records on individuals are standardized and have very little research value, except those about important people. Only summary data about people usually are important for population, sociological, or similar studies. The best example is census records.

Lists of names, such as passenger lists and other records held by NARS, are sometimes important to genealogists.

Records of corporate bodies include docket files of regulatory agencies, some of which may be valuable. As with individuals, records of firms may be valuable in the aggregate only, unless a single firm is historically important.

- **Things.** Informational values about things involve historic structures, naval vessels, roads, patent files, and manmade objects in general. The values derive from the information that records contain about the things themselves, not from information about what happens to the things. Examples drawn from the holdings of NARS are ship plan files, patent files, and public buildings files.

Records on buildings are archivally important only if the buildings themselves are important. The buildings may be significant because they are identified with important people or with architectural ingenuity.

- **Places.** Cartographic information about places refers to specific individual localities at various levels - areas, regions, States, counties, minor civil divisions, or other geographic units. The basic information contained in these records generally pertain to the geographic character of a locality or the relationship of cultural and physical landscape features. Records that reveal information about places include maps and charts, aerial photography, remote sensing imagery, still and motion pictures, field survey notebooks, place name decision files, and site location reports. Correspondence, reports, publications, and other written materials may also contain information about the topological, geological, and geographical features of an area as well as its history.
- **Phenomena.** Records relating to phenomena contain data about what happens to persons or things. They describe conditions, activities, events, episodes, and circumstances. The appraiser has little problem when records deal with a specific event, such as a battle, an election, or a natural catastrophe. Records on phenomena largely concern the masses of statistical data on social and economic matters, such as industrial production, population, prices, income, living costs, and the like. Appraisers need to move slowly and deliberately in evaluat-

ing statistical records. They will certainly need to consult with NARS specialists, who will know the interests of social scientists and others in assessing potential research.

What should be done with the raw data on which statistical records are based? For data not machine-readable, four principles should be followed:

a. Raw data do not need to be kept if all possible studies have been made.

b. If the data have not been used promptly, they will probably not be used at all.

c. Once it is determined that data are not usable because of their physical form or because they are statistically faulty, they should be destroyed.

d. Scientific data should be kept if they are essential for further research.

**Appraisal of Machine-Readable Records**

The appraisal of machine-readable records is especially important because they record much of the statistical data affecting phenomena. Appraisal of a machine-readable file is based on

- Subject matter.
- Its potential for statistical analysis.
- Validity of the recorded data.
- Physical condition of the records.

The appraiser must know the techniques of manipulating computerized data. If only limited manipulation is possible, the appraiser must weigh that factor against the nature of the subject matter to determine whether the file has archival value. In addition, the appraiser must be enough of a technician to be able to examine the physical aspects of the file in determining its readability, such as data density, the number of tracks, and the nature of the data groups.

The continuous changes in computer technology present problems. Compatibility between computer modes, systems, and different manufacturers' products needs to be watched. The danger of erasure is always present.

Many historians, economists, and other social scientists produce studies on social and economic subjects through manipulation of data on machine-readable records. By proper programming, they can correlate data in ways not easily possible with paper media. For example, data in census records can be related to income and population distribution.

Records officers need to be familiar with potential uses of machine-readable data and the special problems of appraisal they represent. When machine-readable records are accessioned by NARS, related system documentation is taken also. That documentation, usually in written form, is vital for programming retrieval of data on tapes.

**Cost Factors**

Appraisers may find that the cost of keeping a series of only marginal value would be considerable. In that case, permanent retention must be carefully justified. The appraiser must show that if the series were to be destroyed, important, irreplaceable documentation would be lost.

**Sampling**

Occasionally, samples of a limited number of transaction case files may be preserved using one of two methods. The quantitative approach involves the application of statistical theory or methods and is used to preserve records to illustrate how a particular program was administered. Examples include the preservation of every tenth or hundredth case file or of every case file closed in years ending 0 or 5. The qualitative method requires the use of criteria of significance and is used to preserve records for the kinds of data that were accumulated. Frequently used criteria are listed in item 2 of table 4.

**TABLE 4 - PERMANENT RECORDS APPRAISAL GUIDELINES**

The generic series descriptions listed below illustrate the type of records normally appraised for permanent retention by NARS. Because of the wide variety of records created in the Federal Government and the complex nature of the appraisal process, this list cannot detail every type of series that may be appraised for retention. In addition, the list applies only to current records whose life cycle has been carefully controlled. Somewhat different standards apply to records created in earlier periods of our history when the maintenance and disposition of Federal records were not as closely regulated. Because many important 19th-century records were inadvertently destroyed by fire, flood, and general neglect, routine administrative and housekeeping records are often preserved for this period to show the functions of the Federal Government.

1. **General Subject Files Documenting Substantive Agency Programs**

   Correspondence with other Federal agencies, Members of Congress and congressional committees, the Executive Office of the President, the President, private organizations and individuals, internal agency memorandums, narrative and statistical reports, budget estimates and justifications, and a variety of other records concerning all substantive and distinctive programs of the agency. These series represent the basic system of records documenting the evolution of major policies and procedures and are frequently designated for permanent retention when created at the following levels: secretary; under secretary; deputy secretary; assistant secretary; administrators, chairpersons, commissioners, and

directors of administrations, bureaus, and services within a department; and heads of independent Federal agencies and their chief assistants. When the agency's important programs are not documented in program correspondence maintained at these higher levels, similar records created at lower levels must be designated for preservation. The number of series selected from a given agency will depend on the degree of duplication evidenced by comparisons among files created at the various administrative levels. Where substantial duplication does exist, the file created at the highest level will be chosen. Where little or no duplication exists, series at all levels will be taken and in some cases at levels lower than those indicated above.

2. **Selected Case Files**

Many Federal records are created in the form of case files. These records may include correspondence, memorandums, periodic narrative reports, and similar materials which relate to a specific action, event, person, place, project, or other subject and provide complete documentation of an agency's activities from initiation to conclusion. Although most case file series are disposable at some future date, a complete set occasionally may be designated for permanent retention, particularly when the files have been captured in machine-readable form. More frequently, however, only a portion of a case file series is selected for transfer to the National Archives. Those chosen normally fall under one or more of the following categories. The case:

a. Established a precedent and therefore resulted in a major policy or procedural change;

b. Was involved in extensive litigation;

c. Received widespread attention from the news media;

d. Was widely recognized for its uniqueness by established authorities outside the Government;

e. Was reviewed at length in the agency's annual report to the Congress; or

f. Was selected to document agency procedures rather than to capture information relating to the subject of the individual file.

Categories (a) through (e) establish the exceptional nature of a particular case file while category (f) relates to routine files chosen because they exemplify the policies and procedures of the creating agency. The types of case files selected for permanent retention under the criteria established above include, but are not limited to, research grants awarded for studies; research and development projects; investigative, enforcement, and litigation case files; social service and welfare case files; labor relations case files; case files related to the development of natural resources and the preservation of historic studies; public works case files; and Federal court case files.

**3. Analytical Reports**

Analytical research studies and periodic reports prepared by the agency or by a private organization or individual under contract to the agency or in receipt of a grant from the agency. Studies and reports selected for permanent retention may be statistical, narrative, machine-readable or audiovisual in nature. Regional reports prepared by field offices and forwarded to the agency's headquarters are frequently selected because they contain information relating to ethnic, social, economic, or other aspects of specific geographical locations. Excluded from selection are studies and reports which are published and therefore widely available in public libraries, as well as recurring periodic reports which are summarized on an annual basis. (See item 13 for publications permanently retained.) In some instances, only selected studies and reports are maintained for future research.

**4. Formal Minutes of Boards and Commissions**

Minutes of meetings of boards and commissions of Federal agencies documenting substantive policy and procedural decisions. Frequently, the executive direction of a Federal agency is provided by a board or commission rather than by a single appointed individual. Typically, these agencies are regulatory bodies such as the Federal Trade Commission, but also include organizations such as the Pension Benefit Guaranty Corporation and the Commission of Fine Arts. Minutes may be literal transcriptions or edited summaries. Sound recordings of these meetings should also be preserved.

**5. Records of Internal Agency, Interagency, and Non-Federal Committes**

Minutes, agenda, proposals submitted for review, and final recommendations of meetings of ad hoc committees as well as more formally established councils, conferences (e.g., White House Conferences), and task forces attended by senior agency officers. These meetings may be limited to internal agency personnel or may include representatives from other Federal agencies or even non-Federal groups. Records selected for permanent retention to document interagency meetings will be limited to the agency designated as the group's secretariat. The minutes selected may be summary in nature, verbatim transcripts, or audio or video recordings.

**6. Legal Opinions and Comments on Legislation**

Memorandums prepared by an agency's legal counsel or program officers concerning interpretations of existing laws and regulations or the effects of proposed laws and regulations which govern the agency or which

have a direct effect on its operations. Records selected under this item concern the agency's primary missions and normally exclude general opinions and comments relating to other Federal agencies. Included are formal comments on pending legislation prepared at the request of the Congress or the Office of Management and Budget. Most of these records are permanent when created in offices of general counsels of departments and independent agencies. Excluded are copies of bills, hearings, and statutes held for convenient reference. Similar records maintained below the departmental level may not be archival depending on their content and relationship with records of the departmental counsel.

7. **Evaluations of Internal Operations**

Studies conducted to determine the effectiveness of the procedures adopted to achieve established policy goals. These may include evaluations of both program and administrative operations and may be made by the agency itself (inspectors general) or by outside oversight agencies (General Accounting Office). Only those studies which recommend significant changes in policy or procedural violations are preserved. In addition, a complete record set of studies prepared by oversight agencies are designated for preservation in the creating agency. All other copies are disposable.

8. **Formal Directives, Procedural Issuances, and Operating Manuals**

Formal directives distributed as orders, circulars, or in loose-leaf manual form announcing major changes in the agency's policies and procedures. Normally these are issued with the authority of the head of the agency. Extensive procedures are frequently detailed in lengthy operating manuals.

9. **Records on Functional Organization**

a. **Organizational charts and reorganization studies.**
Graphic illustrations which provide a detailed description of the arrangement and administrative structure of the functional units of an agency. Reorganization studies are conducted to design an efficient organizational framework most suited to carrying out the agency's programs and include materials such as final recommendations, proposals, and staff evaluations. These files also contain administrative maps that show regional boundaries and headquarters of decentralized agencies or that show the geographic extent or limits of an agency's programs and projects.

b. **Functional statements.**
Formally prepared descriptions of the responsibilities assigned to the senior executive officers of an agency at the division level and

above. If the functional statements are printed in the Code of Federal Regulations, they are not designated for preservation as a separate series.

**10. Briefing Materials**

Statistical and narrative reports and other summary materials prepared for briefings of recently appointed heads of agencies and their senior advisors to inform them of the current status of the agency. In addition, briefing books are occasionally prepared to inform an agency head of the current status of a major issue confronting the agency or in preparation for hearings, press conferences, or major addresses.

**11. Public Relations Records**

a. **Speeches, addresses, and comments.**
Remarks made at formal ceremonies and during interview by heads of agencies or their senior assistants concerning the programs of their agencies. The speeches and addresses may be presented to executives from other Federal agencies, representatives of State and local governments, or private groups, such as college and university students, business associations, and cultural organizations. Interviews may be granted to radio, television, or printed news media commentators. The format selected may be paper, audio or video tape, machine-readable tape or discs, or motion picture film.

b. **News releases.**
One copy of each prepared statement or announcement issued for distribution to the news media. News releases announce events such as the adoption of new agency programs, termination of old programs, major shifts in policy, and changes in senior agency personnel and may be a textual record such as a formal press release or nontextual record, such as film and video or sound recordings.

**12. Agency Histories and Selected Background Materials**

Narrative agency histories including oral history projects prepared by agency historians or public affairs officers or by private historians under contract to the agency. Some background materials (such as interviews with past and present personnel) generated during the research stage may also be selected for permanent retention. Excluded are electrostatic copies of agency documents made by the researcher for convenient reference.

**13. Publications**

Formally prepared publications printed by the Government Printing Office, the National Technical Information Service, or the agency itself.

Examples of such publications include annual reports to the Congress; studies conducted by the agency or under contract for the agency; procedural brochures, pamphlets, and handbooks distributed for guidance to other Federal agencies, State and local governments, and private organizations and citizens; instructional and educational materials on audiovisual formats (audio or video recordings, motion picture, filmstrips and slide-tape productions); maps; and film productions and television and radio programs prepared to furnish information on agency policies or promote agency programs and operations. The availability of reference copies of audiovisual items in non-Government depositories does not exclude retaining the original production elements required to ensure the preservation of the audiovisual items.

**14. Visual, Audio, and Graphic Materials**

Agency-originated motion picture film, still photography, sound and video recordings, cartographic materials, or architectural drawings created to record substantive events or information that cannot be or normally are not recorded in written form. Examples of these materials are instantaneous recordings or photographic coverage of significant scientific or technological phenomena and significant nonrecurring events, such as combat operations, lunar explorations, and extemporaneous occurrences, discussions, and interviews; maps recording topographic information for specific geographic areas; and architectural/engineering drawings recording the building program of individual Federal agencies.

**15. Scientific and Technical Data**

Data resulting from observations of natural events or phenomena or from controlled laboratory or field experiments. These data generally are created at project or operating levels rather than at administrative levels. The data may be recorded in either human-readable or machine-readable format and be found in laboratory notebooks, completed forms, tabulations and computations, graphs, microforms, or machine-readable files. Scientific and technical data are selected for permanent preservation if they are unique, usable, and important. If these data are accurate, comprehensive, and complete, if they can and are likely to be applied to wide variety of research problems, then they can also be considered to have passed the test of usability. Data which can be recreated because they document repeatable activities may also be considered both unique and usable if they constitute a definitive, critical, or standard reference data set. The cost of data collection is one, but not the only, measure of its importance. In assessing the importance of any set of data, consideration should be given to its historical as well as its scientific significance.

16. **Socio-economic Micro-Level Data**

Micro-level data collected for input into periodic and one-time studies and statistical reports including information filed to comply with Government regulations. The information may cover such subjects as economic and tax information, health care, demographic trends, education, discrimination, and other comparable social science areas. Although agency reports and studies, briefing materials, and official releases frequently summarize these data, the micro-level information, usually in machine-readable form, is of permanent value. Obviously, the data must be usable in their raw state if they have not been converted to a machine-readable form.

# APPENDIX A

# APPRAISAL FORMS AND PROCEDURES

## A.3

## APPRAISAL FORM

## AND

## RELATED PROCEDURES

## DEVELOPED BY THE

## MACHINE READABLE ARCHIVES DIVISION,

## PUBLIC ARCHIVES OF CANADA

Public Archives Canada / Archives publiques Canada

# APPRAISAL FORM – FORMULE D'ÉVALUATION

1. Transferring Dept./Agency/Organization – Ministère/Organisme responsable

Machine Readable Data File Sequential No. – N° sequentiel du fichier des archives ordinolingues.

2. Descriptive File Title – Titre descriptif du fichier

3. Data are readable?
Les données sont-elles lisibles?
☐ Yes Oui ☐ No Non

Data are documented?
Sont-elles documentées?
☐ Yes Oui ☐ No Non

Data Source?
Source?
☐ Government Gouvernement
☐ Private Secteur privé

Data Type? – Types de données
☐ Survey Enquête
☐ Data Base Base de données
☐ Graphic Graphique
☐ Other (specify) Autre (préciser)
☐ Econometric Econométriques
☐ Bibliographic Bibliographiques
☐ Cartographic Cartographiques

4. Content Analysis – Analyse du contenu

5. Technical Analysis – Analyse technique

6. Evaluation – Évaluation

7. Appraisal Decision – Décision

☐ Acceptable file Dossier acceptable
☐ Rejected file Dossier rejeté

Archivist – Archiviste

Supervisor – Superviseur

Director, MRA – Directeur, DAO

Date

ARC - 44 (82/06)

N O T E

IL N'Y A PAS DE PAGE 136 - ERREUR DE PAGINATION

(SAUTE A 137)

## APPRAISAL FORM

### - Procedures -

1. **Transferring Dept./Agency/Organization**

Name the department, agency, or organization from which the machine readable data file will or could be potentially acquired.

**Machine Readable Data File Sequential Inventory No.**

Indicate the appropriate Machine Readable Data Inventory number for the file. (A Machine Readable Data Inventory form, ARC-890, must be completed.)

2. **Descriptive File Title**

Give the full descriptive title of the file, including the dates to which the data pertain.

3. **Data are Readable?**

Indicate if the data, as they are stored on the particular device (tape, disk, punched card, etc.) managed by the transferring agency, can be successfully read. If "no," check off "rejected file" in block 7 and submit the form for approval.

**Data are Documented?**

Indicate if sufficient documentation accompanies the data to permit an archivist to compile a documentation package which will adequately meet the needs of researchers interested in using the data. If "no," check off "rejected file" in block 7 and submit the form for approval.

**Data Source?**

Indicate if the data are the property of a government institution, or a private institution or individual.

**Data Type?**

Indicate the type of data by checking the appropriate box.

4. **Content Analysis**

Describe, in narrative form, the purpose and general use of the data as well as the role with respect to the overall functions of the programme(s) involved.

Describe, in narrative form, the research value of the data by indicating the degree to which they have research value. As all data are considered to have at least some value, it will not be possible to record simply "no value." It will be possible, however, through the justification provided in the narrative, to suggest that data may have extremely limited archival value.

As well as the research value of the data, consideration must also be given to the level of aggregation of the data, their linkage potential to other data files or other series of data, and the degree to which the removal of identifiable data fields (that is, anonymization) could limit the usefulness of the file.

5. **Technical Analysis**

Describe, in narrative form, those technical attributes which might affect the appraisal decision. Provide details on the data type, size, complexity, associated hardware and software features, updating characteristics, amount and complexity of documentation, etc.

Duplicate records, if they exist, must be considered if it is determined that they are available, have a more desirable arrangement, and could be less expensive to preserve.

6. **Evaluation**

Based on the results of the "Content Analysis" as weighed against the "Technical Analysis," provide a narrative describing the justification used for either accepting or rejecting the file. This section must be completed in consultation with the supervisor.

In many cases, the evaluation should be fairly straightforward. A file, for instance, may have such high research value and absorb such little cost in terms of acquisition and archival management, that a recommendation to "accept" can be easily justified. Similarly, a file

that has extremely limited research value and yet could potentially absorb substantial costs in terms of acquisition and archival management, could easily be recommended for rejection.

In other cases, however, the justification to "accept" or "reject" could be more difficult to develop. Weighing the research value of a file against the costs of its acquisition and archival management will demand the expertise of at least the archivist and the supervisor and, in certain cases, the Divisional Management Committee. The supervisor will be responsible for ensuring that the justification is in accordance with divisional priorities and collection policies. The justification will also depend upon the archivist's knowledge of the subject area and the department/organization involved. In many cases it will require consultation with departmental/organizational users of the file, as well as any research groups which might have a potential interest in the file.

In most cases, a file appraised as "acceptable" should be acquired immediately. In some cases, however, the decision to acquire may be postponed. The operational life of the data, for example, and the willingness of the department/organization to assist in the servicing of the data, could affect the timing of acquisition and even the archival format.

7. Appraisal Decision

Based on the Evaluation, indicate if the machine readable data file should be accepted or rejected.

8. Acquisition Plan

This section outlines the steps that will be followed to acquire the file from the donor, assuming that the decision to "accept" the file has been formally approved. The format of the data, size, number of tapes required, programs required, computer centres involved, and the general roles of both the PAC/MRA and the donor will be described. The amount and types of documentation and their method of acquisition (photocopying, microfilm, etc.) should also be indicated.

If the file is not to be acquired immediately, indicate the possible acquisition data and describe a possible acquisition strategy.

Calculate the approximate costs of the acquisition (tape copying, program preparation, etc.) and indicate in the space provided. The person/costs or salary dollars will be provided by the supervisor, based on the amount of person/time identified.

9. **Processing Plan**

This section describes a suggested processing strategy and outlines the steps that will be taken to process the acquired file into an archival format. Verification checks (label, dump, frequencies) as well as the degree of verification (partial dump only, all data items verified, etc.) should be described. Documentation preparation and arrangement, including the degree to which the documentation will be converted to an archival standard, should also be indicated.

The final archival format of the data should be generally described in terms of size (number of characters, number of records, etc.), arrangement (sequential, system dump, etc.), complexity, number of tapes required, software dependency (if any), hardware dependency (if any), amount of documentation, etc.

Calculate the approximate cost of processing the file (computer processing, conversion, verification, documentation preparation, etc.) and indicate in the spaces provided. The person/costs will be provided by the supervisor, based on the amount of person/time identified.

10. **Conservation Plan**

This section describes a conservation strategy which will be used to assist conservation staff in preserving the data file for the period of time until the first review date. This section will usually indicate the number of tape rewinds (for example, 10 rewinds over 10 years), the number of tape recopies (for example, 2 recopies over 10 years), and the amount of documentation to be microfilmed (for example, 2 metres in the first year).

Calculate the costs of preservation for the total number of years involved and then divide by the number of years to find the per year cost. The person/costs will be provided by the supervisor, based on the amount of person/time identified.

11. **Access Restrictions**

Based on contacts with appropriate departmental/organizational officials, describe the access restrictions which will be applied to the data and the length of time for which they will be in effect, if applicable. The final access restrictions must be detailed in a separate letter addressed to the Division Director.

12. **Service Plan**

This section outlines a possible servicing strategy that could be followed to service the data for the period of time until the first review date. The most likely type of research request should be described (for example, straight tape copy and documentation, extract, summary, statistical analysis, record retrieval, etc.). Indicate if any conversion of the archival format to a format suitable for research will be required. Describe the number of tapes involved in a typical research request and the amount of documentation which will usually be requested. The number of potential requests per year (estimated) should also be indicated.

If a special arrangement has been made with the donor department/organization or some other agency regarding the servicing of the data, this should be indicated.

Calculate the costs of servicing the data (for example, tape copying, documentation preparation, system file restoration, programming, etc.) for the total number of years involved and then divide by the number of years to find the per year costs. The person/costs will be provided by the supervisor, based on the amount of person/time identified.

13. **Review Date**

All machine readable data files acquired by the Division must be reviewed for continued preservation. "Review Date" refers to the year in which the data should be reappraised. (For a fuller explanation on the approach to be followed in the issuance of a "Review Date", refer to the divisional policy statement and procedure paper on reappraisal.)

14. **Signatures of Approval**

# APPENDIX B

## GLOSSARY OF TERMS

* **ACCESS**

See DIRECT ACCESS and SEQUENTIAL ACCESS.

**ADDRESS**

A name, number, or label that represents a register or a particular location in computer memory or storage.

* **ANALOG**

Pertaining to representation of information by means of continuously variable physical quantities such as electric voltage, as opposed to digital.

* **APPLICATION**

A problem or task to which automated techniques are applied.

**APPLICATION SOFTWARE**

A computer program intended to solve a specific problem or do a job, as distinct from system software which controls the operation of a computer system or performs many functions. See also OPERATING SYSTEM and SYSTEM SOFTWARE.

**ARITHMETIC AND LOGIC UNIT (ALU)**

The portion of the central processing unit in which arithmetic and logical operations are performed. See CENTRAL PROCESSING UNIT (CPU).

**ASCII (AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE)**

The standard character code, widely used to represent characters in binary code. ASCII is one of several codes which assigns a specific pattern of binary digits (bits) to each character. The standard ASCII code has eight bits. ASCII-10 is a standard code for telecommunications.

**ASSEMBLY LANGUAGE**

A computer language used to write programs with symbolic instructions, rather than with binary instructions in machine language that can be understood directly by the CPU. The computer uses a program called an assembler to translate the instructions into its own machine language. Assembly language is a low-level language which allows the programmer to write instructions using terms like "LOAD", "RUN", and "CLEAR" instead of as strings of 0's and 1's.

**AUTOMATIC DATA PROCESSING (ADP)**

Refers to the manipulation of information by a computer in accordance with a set of instructions. It is used synonymously with the term ELECTRONIC DATA PROCESSING (EDP).

**BACK-UP FILE**

A file containing information which duplicates information in other files and which can be used to replace or recreate the original file if its contents are lost or destroyed.

**BCD (BINARY CODED DECIMAL)**

A code for representing decimal numbers with binary digits (bits). BCD is a six-bit code.

* **BINARY**

Pertaining to an object, a condition, an action, or a variable which can have one of two possible values or states.

Pertaining to the number representation system with base two, using the digits 0 and 1.

**BINARY DIGIT**

A numeral in the binary numbering system, either 0 or 1. In electronics, 0 commonly represents absence of a pulse and 1 represents presence of a pulse.

\+ **BINARY NOTATION**

A system in Automatic Data Processing in which numbers are represented by the two digits 0 and 1.

+ **BIT**

In binary notation, one of the two digits 0 and 1. An abbreviation of b(inary dig)it.

**BITS PER INCH**

Abbreviated as b.p.i. It is usually synonymous with characters per inch. A measure of bit density. See DENSITY.

**BLOCK**

A section of recorded information on a magnetic tape or magnetic disc/-disk, separated from other blocks by a small area of non-recorded tape or disc/disk, called an inter-record gap or IRG. One block may contain many logical records, or one logical record may extend over several blocks depending upon the size of the records, the hardware used, and the decisions of the programmer. A block is also called a physical record.

+ **BYTE**

A string of adjacent bits, usually eight, operated upon as a unit in Automatic Data Processing.

The representation of a character in Automatic Data Processing.

**CATHODE RAY TUBE (CRT)**

A device that displays data in visual form by means of controlled electron beams, but generally refers to any television-type screen.

* **CENTRAL PROCESSING UNIT (CPU)**

The part of a computer in which processing takes place. It includes the control and arithmetic units and main storage. Formerly also called main frame, a term now generally used more loosely to qualify a large and sophisticated computer, as opposed to a minicomputer.

+ **CHARACTER**

A letter, digit, or other symbol used in the representation of data.

**CHARACTER CODE**

A code that represents alphanumeric data in binary form. See also ASCII, BCD, and EBCDIC.

* **CODE**

A set of unambiguous rules for representing information or instructions in symbolic form. For example, a code established by a user to represent his/her data so that it may be processed, and a code established by manufacturers or standards organizations for the representation of characters in binary form.

**CODEBOOK**

An explanation of the codes used to represent information in numeric or abbreviated form.

* **COMPILER**

A program which converts instructions from another language into machine code. With minicomputers a similar function is performed by an interpreter.

\+ **COMPUTER**

A machine for processing data under the control of recorded instructions without intervention of a human operator during operation. See also MICROCOMPUTER and MINICOMPUTER.

**CONTROL UNIT**

The portion of the CPU which supervises the overall operation of the computer and other hardware components. The control unit undertakes the retrieval of instructions and data in the proper sequence, the interpretation of each instruction, and regulates the transfer of data from one hardware component to another.

* **CONVERT**

To change the representation of data from one format to another, on the same or a different medium.

* **DATA**

The formal representation of information prepared with a view to it being handled by an automatic process. Especially that which is given in the

formulation of a problem and is distinct from its results, or that which is the object of processing as opposed to the commands that control the process.

**DATA BASE**

A collection of carefully integrated files, usually stored in a central location and made available to several users simultaneously for a variety of applications. Users may have access to all, or only portions, of a data base.

**DATA BASE MANAGEMENT SYSTEM (DBMS)**

A software system used to access and retrieve data that are stored as data elements in a common file, called a data base. The data elements are stored randomly and each data element has an imbedded address system or pointer. The DBMS allows access to the data and combines data elements into records or files to meet user specifications.

**DATA ELEMENT**

A combination of characters or bytes referring to one separate item of information, such as name, address, age, and so on.

* **DATA PROCESSING**

The execution of a systematic set of operations on data, such as sorting or calculation.

**DATA SET**

A computer-readable collection of records related in some way and treated as a unit. Data set is used synonymously with data file and data base.

* **DATA TRANSMISSION**

The transfer of data derived from or intended for automatic processing from remote terminals to a central computer or from one computer system to another.

\+ **DENSITY**

In Automatic Data Processing, the number of bits or characters that can be stored per unit of dimension of a medium. In the case of magnetic tape, this is expressed as bits per inch (bpi).

**DIGITAL**

Refers to the representation of data in discrete (discontinuous) form in distinction from analog which represents data in continuous form.

* **DIRECT ACCESS**

The method of placing data in storage, or retrieving data from storage, by means only of an address and independently of the data most recently placed or retrieved.

**DISC/DISK**

A circular plate with a magnetic coating on both sides. The plate rotates for the storage and retrieval of data by one or more "read/write heads" which transfers the information to and from the computer. The computer-readable information may be recorded on a soft (floppy) or hard (rigid) disc/disk and may be recorded on one or both sides of the disc/disk.

**DISC/DISK DRIVE**

A device that reads or writes data on magnetic discs/disks.

**DISCETTE/DISKETTE**

A storage device having a thin flexible mylar disc/disk protected by a semi-rigid square plastic cover. Discettes/Diskettes are usually 5¼ or 8 inches in diameter. The read/write head of the disc/disk drive touches the surface or the discette/diskette through an open slot and provides direct access to recorded information. Discettes/Diskettes are often referred to as "floppy" discs/disks to distinguish them from their more rigid counterparts.

**DISC/DISK PACK**

A set of magnetic discs/disks which are fixed to a common spindle. They can be physically removed from the handling unit for storage and replaced by other similar sets.

\+ **DOCUMENTATION**

An organized series of descriptive documents explaining the operating system and software necessary to use and maintain a file and the arrangement, content, and coding of the data which it contains.

**EBCDIC (EXTENDED BINARY CODED DECIMAL INTERCHANGE CODE)**

An eight-bit code for representing characters, used with IBM computers and many others.

**ELECTRONIC DATA PROCESSING (EDP)**

Refers to the manipulation of information by an electronic computer. The term is used synonymously with AUTOMATIC DATA PROCESSING (ADP).

**FIELD**

A specific area of a record that is allocated for a particular category of data, usually one data element. In some applications each field is assigned a fixed number of positions (fixed-length fields), while in other applications the length of fields may vary within defined limits (variable-length fields). In some applications, fields may or may not be present in every record (variable-occurrence fields).

* **FILE**

A set of related records capable of being processed as a unit. See also BACK-UP FILE, MACHINE-READABLE DATA FILE, MASTER FILE, and PROCESSING FILE.

**FILE STRUCTURE**

The way in which a particular file is organized. When the record layout consists of one record per entry, then the file structure is rectangular. When the record layout consists of several records per entry, the file structure is hierarchical.

* **FORMAT (Noun)**

The specified description of the content and organization of data. It is also known as layout.

* **FORMAT (Verb)**

To arrange data in a specified structure or format.

**HARDCOPY**

Information printed on paper or other durable surface, such as microfilm. The term is used to distinguish printed information from the temporary image represented on a CRT screen, and from the machine-readable information on a magnetic tape, disc/disk, etc.

+ **HARDWARE**

The physical units making up a computer system as distinct from software.

**HARDWARE DEPENDENT**

Any machine-readable file that requires one particular make or model of computer or the features unique to a certain configuration of computer equipment is hardware dependent (also called machine dependent). Machine-readable records that can be used on any computer are machine or hardware independent.

* **INPUT**

The insertion of data into a computer or one of its parts. An input device is one capable of communicating data to a computer.

**INPUT/OUTPUT (I/O)**

A general term for the peripheral equipment used to communicate with a computer; the data involved in such communications; the medium carrying such data; or the activities of reading or writing such data.

**LABEL**

An identifier which provides information about a file, magnetic tape, or direct access device. There are two types of labels-external and internal. External labels identify the physical medium and are used to locate tapes, discs/disks, etc. Internal labels are written in machine-readable form at the beginning and/or end of a file and provide information which identifies the file or records in the file.

**LANGUAGE**

A defined set of symbols and the rules or conventions governing the ways in which the symbols may be combined for the purpose of meaningful communication. See also ASSEMBLY LANGUAGE, MACHINE LANGUAGE, and PROGRAMMING LANGUAGE.

* **LAYOUT**

See FORMAT.

**LOGICAL RECORD**

A compilation of related data elements referring to one person, place, thing, or event that are treated as a unit. Logical records can have a specified number of characters (fixed-length records) or the number of characters in each record can vary within limits (variable-length records).

**LOGICAL RECORD LENGTH**

The number of characters or bytes in a logical record.

**MACHINE LANGUAGE**

The lowest level language, written in binary code, that a computer understands directly without translation. Most computers are equipped with the capability of using higher level languages which are translated into machine language by programs called assemblers or compilers that are stored in the computer as part of the operating system. See also ASSEMBLY LANGUAGE and PROGRAMMING LANGUAGE.

**MACHINE-READABLE**

Information in a form that can be processed directly by a computer, usually in the form of magnetic or electronic impulses.

**MACHINE-READABLE DATA FILE (MRDF)**

A body of information encoded and formatted in such a way that it requires the use of a data processing machine or computer to be properly interpreted.

* **MAGNETIC TAPE**

Tape with a magnetizable surface on which data can be recorded for storage by selective polarization of the surface.

* **MAGNETIC TAPE DRIVE**

A device that moves magnetic tape past a head for reading and writing.

**MAINFRAME**

The main body of a large computer, in distinction from its peripherals and from minicomputers and microcomputers.

**MASTER FILE**

A relatively permanent machine-readable data file containing an organized, consistent set of records of complete and accurate information.

* **MEDIUM**

Material whose physical characteristics can be changed at various points to record data, such as magnetic tape, punched card, etc.

\+ **MEMORY**

Any device into which data can be recorded and stored and from which they can be retrieved.

* **MERGE**

To combine items from two or more similarly ordered sets into one set or file arranged in the same order.

* **MICROCOMPUTER**

A computer in its simplest form based on a microprocessor and limited main memory and lacking storage. Originally dedicated to a single function, such computers are now capable of handling several tasks.

**MICROLEVEL DATA**

Data at the lowest level of aggregation possible, generally pertaining to each case, event, transaction, etc.

\+ **MICROPROCESSOR**

A central processing unit contained on a single silicon chip.

* **MINICOMPUTER**

A general purpose computer characterized by its relatively low cost, small size, and ability to operate in normal office environmental conditions. It may be linked as a terminal to a mainframe.

* **MULTIPROCESSOR**

A computer employing two or more processors operating under integrated control and sharing storage.

* **OPERATING SYSTEM**

Software which controls the execution of computer programs. This software is specific to a particular type of machine and, besides control, generally provides for compiling, debugging, scheduling, and file handling.

* **OUTPUT**

Data delivered from a computer. Pertaining to a part of a computer or to a peripheral capable of performing such delivery. Also pertaining to the execution of such delivery. See also INPUT/OUTPUT.

* **PACKAGE**

Software designed to carry out similar functions for a variety of users, each of them specifying any adaptations required in a manner set out in the documentation forming part of the package.

* **PERIPHERAL**

Pertaining to a device separate from the central processing unit which can be connected to and controlled by that unit.

**PHYSICAL RECORD**

A record treated as a unit because of its physical form (for example, a block). A collection of data defined in terms of physical parameters, rather than logical content.

**PRINTOUT**

Paper documents containing an eye-readable version of all, or selected portions, of the information in a machine-readable data file.

**PROCESSING FILE**

A relatively temporary file used to generate, correct, reorganize, or derive output from a master file or to store data used to update a master file. Processing files are also called transaction files and production files.

* **PROGRAM**

A series of instructions or statements in a form acceptable to a computer prepared in order to achieve a particular result.

* **PROGRAMMING**

The process of designing, writing, and testing a program.

* **PROGRAMMING LANGUAGE**

A set of representations, conventions, and rules to convey instructions to a computer. A language which permits instructions to be written in a somewhat more natural manner and in a more compressed form is known as a high-level language. Such languages are relatively machine independent.

* **PUNCHED CARD**

A card that can be punched with a pattern of holes to represent data. Hole positions are most commonly arranged in 80 columns and 12 rows. Other cards have 96 columns and 4 rows.

**RAW DATA**

Data which have not been edited or fully processed, but sometimes is used synonymously with microlevel data. See also MICROLEVEL DATA.

* **RECORD**

A collection of related items of data treated as a unit. The basic component of a file.

**RECORD FORMAT**

The arrangement or layout of records on a data carrier. The arrangement may be fixed or variable.

**RECORD LAYOUT**

A diagram or list showing the length of each field, the item of information in each field, and the arrangement of fields in the file. A record layout is also called a file layout.

**REGISTER**

A device used to store a specific amount of data inside the memory of a computer, usually long enough to complete a specific job. Examples of registers are address registers, index registers, and instruction registers.

**SAMPLE**

A subset of a population.

**SAMPLE DESIGN**

The selection of a part (sample) from the whole (population) in order to make inferences about the whole.

**SAMPLE SIZE**

The numerical count of the sample (or subset of the population). The count can be in terms of cases, observations, respondent, etc.

* **SEQUENTIAL ACCESS**

The method of placing data in storage, or retrieving data from storage, in which the data are arranged in sequence, such as on magnetic tape.

\+ **SOFTWARE**

A set of computer programs, procedures, and associated documentation concerned with the operation of a data processing system, as distinct from hardware.

**SOFTWARE DEPENDENT**

Any machine-readable data file that requires a particular computer program, software package, system software, or operating system in order to access, retrieve, or process data in the file.

**STORAGE MEDIA**

The materials on which data are written and stored. Examples include punched cards, magnetic tape, discs/disks, discettes/diskettes, and drums.

**SYSTEM SOFTWARE**

Programs that belong to an entire system rather than a single user and that usually perform or support a function.

**TAPE DRIVE**

A device used to record, read, and erase data on magnetic tape. The tape drive controls the movement of magnetic tape and is used to pass the tape under the read/write assembly and rewind the tape.

* **TAPE LIBRARY**

A store for magnetic tapes containing data which have continuing value for the creating computer centre including associated personnel and retrieval systems.

* **TELEPROCESSING**

Data processing with the use of telecommunications facilities.

**TEXTUAL**

Refers to information written in human-readable form.

* **TRANSCRIBE**

To transfer data from one medium to another, performing conversions as required by the new medium. In a tape library or machine-readable archives, to copy from one magnetic tape to another in the interests of the preservation of data.

* **UPDATE**

To introduce new data to a file by additions of new records or fields and-/or deletion and amendment of old ones, without changing their structure.

**UTILITY PROGRAM**

A program made available by the operating system to save programmers the work of writing their own programs for often needed tasks.

**VARIABLE**

A name for a quantity. A variable can be thought of as a box into which a value may be written. Usually each data element is a variable. Each question in a survey may also be a variable.

**VARIABLE LIST**

A sequential set of variables whose values constitute the cases in a file. Each variable has a mnemonic; the sequence of mnemonics represents the variable list.

**SOURCES**

+ Computer terms which will appear in the forthcoming ICA Multilingual Glossary of Archival Terminology, compiled by Peter Walne.

* Definitions extracted from a planned publication describing Elementary Terms in Archival Automation, compiled by members of the ICA's Automation Committee.

  Sue A. Dodd, Data Element Dictionary for Describing Machine-Readable Data Files (MRDF), June 1983 edition. Prepared for a two-day workshop on machine-readable records presented prior to the 1983 annual conference of the Society of American Archivists held in Minneapolis, Minnesota.

  Margaret L. Hedstrom, Archives and Manuscripts: Machine-Readable Records. Basic Manual Series, Society of American Archivists, Chicago, scheduled for publication in 1984.

# BIBLIOGRAPHY


Arad, A. and Olsen, M.E. An Introduction to Archival Automation. Koblenz, Federal Republic of Germany: Committee on Automation, International Council on Archives, January 1981.

Bell, Lionel. The Archival Implications of Machine-Readable Records. Washington, D.C.: VIII International Congress on Archives, 1976.

Bell, Lionel and Roper, Michael (eds.). Proceedings of an International Seminar on Automatic Data Processing in Archives. London: HMSO, 1975.

Boque, Allan G. "The Historian and Social Science Data Archives in the United States," Library Trends, Volume 25, Number 4, 1977, pages 847-866.

Brichford, M.J. Archives and Manuscripts: Appraisal and Accessioning. Chicago: Society of American Archivists, 1977.

Brown, Cally and Taylor, Marcia. "Who Owns Contract and Grant Data in the U.K. and Who Can Use It?," IASSIST Newsletter, Volume 6, Number 3, Summer 1982, pages 9-15.

Brown, Thomas E. "Appraisal of Machine-Readable Records," an unpublished workbook. Washington, D.C.: National Archives and Records Service, July 1981.

Brown, Thomas E. The Impact of the Federal Use of Modern Technology on Appraisal. Washington, D.C.: An internal report prepared for the Appraisal and Disposition Task Force of the National Archives and Records Service, March 1983.

Brown, Thomas E. "Who Owns Contract and Grant Data and Who Can Use It? A Look at the U.S.A.," IASSIST Newsletter, Volume 6, Number 3, Summer 1982, pages 4-8.

Clubb, Jerome M. "Quantification and the 'New' History: A Review Essay," The American Archivist, Volume 37, Number 1, January 1974, pages 15-25.

Clubb, Jerome M. and Austin, Eric. Computers in History and Political Science. White Plains, New York: IBM, 1972.

Cook, Michael. Archives and the Computer, particularly Chapter 4. Toronto: Butterworth & Co. (Canada) Ltd., 1980.

Dodd, Sue A. Cataloging Machine-Readable Data Files. Chicago: American Library Association, 1982.

Dodd, Sue A. "Data Element Dictionary for Describing Machine-Readable Data File (MRDF)," June 1983 edition, prepared for a two-day workshop on machine-readable records presented prior to the 1983 annual conference of the Society of American Archivists held in Minneapolis, Minnesota.

Dollar, Charles M. "Appraising Machine-Readable Records," The American Archivist, Volume 41, Number 4, October 1978, pages 423-430.

Dollar, Charles M. "Computers, the National Archives, and Researchers," Prologue, Volume 8, Number 1, Spring 1976, pages 29-34.

Dollar, Charles M. "Documentation of Machine-Readable Records and Research: A Historian's View," Prologue, Volume 3, Number 1, Spring 1971, pages 27-31.

Fishbein, Meyer H. "Appraisal of Twentieth Century Records for Historical Use," Illinois Libraries, Volume 52, Number 2, 1970, pages 154-162.

Fishbein, Meyer H. "Appraising Information in Machine Language Form," The American Archivist, Volume 35, Number 1, January 1972, pages 35-43.

Fishbein, Meyer H. Guidelines for Administering Machine-Readable Archives. Washington, D.C.: Committee on Automation, International Council on Archives, November 1980.

Fishbein, Meyer H. "The Evidential Value of Contextual Records: An Early Precedent," The American Archivist, Volume 45, Number 2, Spring 1982, pages 189-190.

Flood, Roderick. "Quantitative History: Evolution of Methods and Techniques," Journal of the Society of Archivists, Volume 5, Number 7, 1977, pages 407-417.

Gavrel, Katharine and McDonald, John. "Appraisal Guidelines," unpublished appraisal package used in the Machine Readable Archives Division, Public Archives of Canada, 1980.

Geda, C.L., Austin, C.W., and Blouin, Jr., F.X. (eds.). Proceedings of a Conference on Archival Management of Machine-Readable Records, Held at the Bentley Library, the University of Michigan, February 1979. Chicago: Society of American Archivists, 1980.

Geda, Carolyn L. "Social Science Data Archives," The American Archivist, Volume 42, Number 2, April 1979, pages 158-166.

General Services Administration, National Archives and Records Service. General Records Schedule 20, Machine-Readable Records, FPMR 101-11.4, February 16, 1977.

Hays, Samuel P. "The Use of Archives for Historical Statistical Inquiry," Prologue, Volume 1, Number 2, Fall 1969, pages 7-15.

Hedstrom, Margaret L. Archives and Manuscripts: Machine-Readable Records. Chicago: Society of American Archivists, scheduled for publication in 1984.

Hedstrom, Margaret L. "Basic Computer Concepts Workshop," package prepared for the 45th annual meeting of the Society of American Archivists held in Berkeley, California, September 1981.

Hickerson, H. Thomas. Archives and Manuscripts: An Introduction to Automated Access. Chicago: Society of American Archivists, 1981.

Hull, Felix. The Use of Sampling Techniques in the Retention of Records: A RAMP Study with Guidelines. Paris: General Information Program and UNISIST, UNESCO, 1981.

Kesner, Richard M., compiler and editor. Information Management, Machine-Readable Records, and Administration: An Annotated Bibliography, particularly Chapter 4. Chicago: Society of American Archivists, 1983.

Knoppers, Jake V. Th. Managing the Electronic Revolution, Archiving the Electronic Heritage: A Position Paper on Issues and Problems in EDP Records/Data Management. Ottawa: Prepared under contract for the Machine Readable Archives Division, Public Archives of Canada, December 1981.

Knoppers, Jake V. Th. Report on Archival Sampling Strategy and Related Issues. Ottawa: Prepared under contract for the Machine Readable Archives Division, Public Archives of Canada, October 1983.

Knoppers, Jake V. Th. Report on Sovereignty Aspects of Transborder Data Flows. Ottawa: Prepared under contract for the Canadian Department of Communications, September 1982.

Knoppers, Jake V. Th. Towards a Canadian Electronic Cultural Heritage/Vers un patrimoine Canadien informatisé. Ottawa: A discussion paper prepared under contract for the Department of Communications and the Public Archives of Canada, May 1983.

McDonald, John F. The MRA and System 2000 Data Bases. Ottawa: A report of a special project undertaken in the Machine Readable Archives Division, Public Archives of Canada, April 1980.

Rapport, Leonard. "No Grandfather Clause: Reappraising Accessioned Records," The American Archivist, Volume 44, Number 2, Spring 1981, pages 143-150.

Schellenberg, Theodore R. The Appraisal of Modern Records. Washington: National Archives, 1956.

Selbee, Kevin. Report on Data File Anonymization. Ottawa: Prepared as a result of a personal service contract established with the Machine Readable Archives Division, Public Archives of Canada, August 1979.

State Historical Society of Wisconsin. Archival Preservation of Machine-Readable Public Records: The Final Report of the Wisconsin Survey of Machine-Readable Public Records. Madison, Wisconsin: The State Historical Society of Wisconsin, 1981.

Thibodeau, Kenneth. "Machine Readable Archives and Future History," Computers and the Humanities, Volume 10, Number 2, 1976, pages 89-92.

White, Howard D. (ed.). "Machine-Readable Social Science Data," Drexel Library Quarterly, Volume 13, Number 1, 1977.

1. Unesco. General Information Programme. Expert Consultation on the Development of a Records and Archives Management Programme (RAMP) within the framework of the General Information Programme, 14-16 May 1979. Paris, Working Document (PGI-79/WS/1). Paris, Unesco, 1979. 19 p. Available also in French.

2. Unesco. General Information Programme. Expert Consultation on the Development of a Records and Archives Management Programme (RAMP) within the framework of the General Information Programme, 14-16 May 1979. Paris, Final Report (PGI-79/WS/II). Paris, Unesco, 1979. 36 p. Available also in French.

3. Manning, Raymond, Gilberte Pérotin and Sven Welander, comps and eds. Guide to the Archives of International Organizations. Part I. The United Nations System. Preliminary version (PGI-79/WS/7). Paris, 1979. 301 p.

4. Cook, Michael. The Education and Training of Archivists - Status Report of Archival Training Programmes and Assessment of Manpower Needs (PGI-79/CONF.604/COL.2). Paris, Unesco, 1979. 71 p. Available also in French.

5. Delmas, Bruno. The Training of Archivists - Analysis of the Study Programme of Different Countries and Thoughts on the Possibilities of Harmonization (PGI-79/CONF.604/COL.1). Paris, Unesco 1979. 75 p. Available also in French.

6. Unesco. Division of the General Information Programme. Meeting of Experts on the Harmonization of Archival Training Programmes, 26-30 November, Paris, 1979, Final Report (PGI-79/CONF.604/COL.7). Paris, Unesco, 1980. 18 p. Available also in French.

7. Roper, Michael. Democratic Republic of the Sudan: Establishment of a Technical Training Centre in Archival Restoration and Reprography (FMR/PGI-80/180). Paris, Unesco 1980, 31 p.

8. Kecskeméti, Charles and Evert Van Laar. Model Bilateral and Multilateral Agreements and Conventions Concerning the Transfer of Archives (PGI-81/WS/3) Paris, Unesco, 1981. 34 p. Available also in Arabic, French, Russian and Spanish.

9. Silva, G.P.S.H. de. A Survey of Archives and Manuscripts Relating to Sri Lanka and Located in Major London Repositories (PCI-81/WS/4). Paris, Unesco, 1981, 100 p.

10. Borsá, Iván. Feasibility Study on the Creation of an Internationally Financed and Managed Microfilm Assistance Fund to Facilitate the Solution of Problems involved in the International Transfer of Archives and in Obtaining Access to Sources of National History Located in Foreign Archives (PGI-81/WS/7). Paris, Unesco, 1981. 31 p. Available also in Arabic, French, Russian and Spanish.

11. White, Brenda. Archives Journals: A Study of their Coverage by Primary and Secondary Sources. (RAMP Studies and Guidelines). (PGI-81/WS/10). Paris, Unesco, 1981. 72 p. Available also in French.

12. Pieyns, Jean. Feasibility Study of a Data Base on National Historical Sources in Foreign Repositories (PGI-81/WS/24). Paris, Unesco, 1981. 66 p. Available also in French.

13. Weill, Georges. The Admissibility of Microforms as Evidence: A RAMP Study (PGI-81/WS/25). Paris, Unesco, 1982. 84 p. Available also in French and Spanish.

14. Hull, Felix. The Use of Sampling Techniques in the Retention of Records: A RAMP Study with Guidelines (PGI-81/WS/26). Paris, Unesco, 1981. 64 p. Available also in French and Spanish.

15. Cortés Alonso Vicenta. Perú: Sistema Nacional de Archivos y Gestión de Documentos: RAMP Proyecto Piloto (FM/PGI-81/110). Paris, Unesco, 1981 56 p.

16. Crespo, Carmen. Republic of Argentina: Development of a Regional Demonstration and Training Centre at the School for Archivists, University of Cordoba (FMR/PGI-81/116 E). Paris, Unesco, 1981. 28 p. Available also in Spanish.

17. Ricks, Artel. Republic of the Philippines: RAMP pilot project for the establishment of a regional archives and records centre. (FMR/PGI-81/158). Paris, Unesco, 1981, 49 p.

18. Evans, Frank B. The Republic of Cyprus: Development of an archival and records management programme (FMR/PGI-81/166). Paris, Unesco, 1981. 64 p.

19. Unesco. General Information Programme. Survey of Archival and Records Management Systems and Services 1982 (PGI-82/WS/3). Paris, Unesco, 1982. Available also in French.

20. Rhoads, James B. The Applicability of UNISIST Guidelines and ISO International Standards to Archives Administration and Records Management: A RAMP Study (PGI-82/WS/4). Paris, Unesco 1982. 95 p. Available also in French and Spanish.

21. Unesco. Division of the General Information Programme. Second Expert Consultation on RAMP (RAMP II) Berlin (West), 9-11 June 1982. Working Document (PGI-82/WS/6). Paris, Unesco, 1982. 31 p.

22. White, Brenda. Directory of Audio-Visual Materials for Use in Records Management and Archives Administration Training (PGI-82/WS/8). Paris, Unesco, 1982. 71 p.

23. Tirmizi, S.A.I. Guide to Records Relating to Science and Technology in the National Archives of India: A RAMP Study (PGI-82/WS/12). Paris, Unesco, 1982. 84 p.

24. Cook, Michael. Guidelines for Curriculum Development in Records Management and the Administrations of Modern Archives: A RAMP Study (PGI-82/WS/16) Paris, Unesco, 1982. 74 p.

25. Unesco. Division of the General Information Programme. Second Expert Consulation on RAMP (RAMP II) Berlin (West), 9-11 June 1982. Final Report (PGI-82/WS/24). Paris, Unesco, 1982. 54 p. Available also in French and Spanish.

26. Evans, Frank B. Malaysia: Development of the Archives and Records Management Programme (FMR/PGI-82/110). Paris, Unesco, 1982. 54 p.

27. Ricks, Artel. Philippines: RAMP Pilot Project for the Establishment of a Regional Archives and Records Centre (Report No. 2). (FMR/PGI-82/161). Paris, Unesco, 1982. 24 p.

28. Evans, Frank B. Writings on Archives published by and with the assistance of Unesco: A RAMP Study (PGI-83/WS/5). Paris, Unesco, 1983. 33 p.

29. Evans, Frank B. and Eric Ketelaar. A Guide for Surveying Archival and Records Management Systems and Services: a RAMP Study (PGI-83/WS/6). Paris, Unesco, 1983. 30 p. Available also in French and Spanish.

30. Hildesheimer, Françoise. Guidelines for the Preparation of General Guides to National Archives: A RAMP Study (PGI-83/WS/9). Paris, Unesco, 1983. 67 p. Available also in French.

31. Kula, Sam. The Archival Appraisal of Moving Images: A Ramp Study with Guidelines (PGI-83/WS/18). Paris, Unesco, 1983. 130 p.

32. Moideen, P.S.M. A Survey of Archives Relating to India and Located in Major Repositories in France and Great Britain (PGI-83/WS/19). Paris, Unesco, 1983. 72 p.

33. Duchein, Michel. Obstacles to the Access, Use and Transfer of Information from Archives: A RAMP Study (PGI-83/WS/20). Paris, Unesco, 1983. 80 p. Available also in French.

34. Rhoads, James B. The Role of Archives and Records Management in National Information Systems: A RAMP Study (PGI-83/WS/21). Paris, Unesco, 1983. 56 p. Available also in French.

35. Hendriks, Klaus B. The Preservation and Restoration of Photographic Materials in Archives and Libraries: A RAMP Study with Guidelines (PGI-84/WS/1). Paris, Unesco, 1984. 121 p.

36. Stark, Marie C. Development of Records Management and Archives Services within United Nations Agencies (PGI-83/WS/26). Paris, Unesco, 1983. 215 p.

37. Kathpalis, I.P. A model curriculum for the training of specialists in document preservation and restoration: A RAMP Study with guidelines (PGI-84/WS/2). Paris, Unesco, 1984. 27 p. Available also in French and Spanish.

38. Seton, Rosemary L. The preservation and administration of private archives: A RAMP Study. (PGI-84/WS/6). Paris, Unesco, 1984. 85 p.

Copies of the above studies and reports may be obtained without charge, to the extent that they are still in print, by writing to:

Division of the General Information Programme,
Documentation Centre,
7, place de Fontenoy,
75700, Paris, France.